PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12N 5/10, 15/12, C07K 14/00, 16/18, G01N 33/53

A1

(11) International Publication Number: WO 96/31625

(43) International Publication Date: 10 October 1996 (10.10.96)

(21) International Application Number: PCT/US96/04454

(22) International Filing Date: 4 April 1996 (04.04.96)

(30) Priority Data:

417,872	7 April 1995 (07.04.95)	US
630,915	3 April 1996 (03.04.96)	US

(81) Designated States: AL, AM, AU, AZ, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LS, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(71) Applicants: CYTOGEN CORPORATION [US/US]; 600 College Road East, Princeton, NJ 08540 (US). UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL [US/US]; Office of Technology Development, CB#4100, 302 Bynum Hall, Chapel Hill, NC 27599-4105 (US).

(72) Inventors: SPARKS, Andrew, B.; 201 Blue Ridge Road, Carrboro, NC 27510 (US). HOFFMAN, Noah; 5001 Manning Drive, Greensboro, NC 27410 (US). KAY, Brian, K.; 18 Wysteria Way, Chapel Hill, NC 27514 (US). FOWLKES, Dana, M.; 2013 Damascus Church Drive, Chapel Hill, NC 27516 (US). MCCONNELL, Stephen, J.; 10211 Camino Ruiz #52, San Diego, CA 92126 (US).

(74) Agents: MISROCK, S., Leslie et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(54) Title: POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME

(57) Abstract

Novel polypeptides having functional domains of interest are described, along with DNA sequences that encode the same. A method of identifying these polypeptides by means of a sequence-independent (that is, independent of the primary sequence of the polypeptide sought), recognition unit-based functional screen is also disclosed. Various applications of the method and of the polypeptides identified are described, including their use in assay kits for drug discovery, modification, and refinement.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME

This application is a continuation-in-part of co-pending U.S. Patent Application Serial No. 08/417,872 filed April 7, 1995, the entire contents of which are incorporated herein by reference.

1. Introduction

The present invention is directed to polypeptides having a functional domain of interest or functional equivalents thereof. Methods of identifying these polypeptides are described, along with various methods of their use, including but not limited to targeted drug discovery.

2. Background of the Invention

Combinatorial libraries represent exciting new tools in basic science research and drug design. It is possible through synthetic chemistry or molecular biology to generate libraries of complex polymers, with many subunit permutations. There are many guises to these libraries: random peptides, which can be synthesized on plastic pins (Geysen et al., 1987, J. Immunol. Meth. 102:259-274), beads (Lam et al., 1991, Nature 354:82-84) or in a soluble form (Houghten et al., 1991, Nature 354:84-86) or expressed on the surface of viral particles (Cwirla et al., 1990, Proc. Natl. Acad. Sci. USA 87:6378-6382; Kay et al., 1993, Gene 128:59-65; Scott and Smith, 1990, Science 249:386-390); nucleic acids (Ellington and Szostak, 1990, Nature 346:818-822; Gao et al., 1994, Proc. Natl. Acad. Sci. USA 91:11207-11211; Tuerk and Gold, 1990, Science 249:505-510); and small organic molecules (Gordon et al., 1994, J. Med. Chem. 37:1385-1401). These libraries are very useful in mapping protein-protein interactions and discovering drugs.

Phage display has become a powerful method for screening populations of peptides, mutagenized proteins, and

cDNAs for members that have affinity to target molecules of interest. It is possible to generate 10^8-10^9 different recombinants from which one or more clones can be selected with affinity to antigens, antibodies, cell surface receptors, protein chaperones, DNA, metal ions, *etc.* Screening libraries is versatile because the displayed elements are expressed on the surface of the virus as capsid-fusion proteins. The most important consequence of this arrangement is that there is a physical linkage between phenotype and genotype. There are several other advantages as well: 1) virus particles which have been isolated from libraries by affinity selection can be regenerated by simple bacterial infection, and 2) the primary structure of the displayed binding peptide or protein can be easily deduced by DNA sequencing of the cloned segment in the viral genome.

Combinatorial peptide libraries have been expressed in bacteriophage. Synthetic oligonucleotides, fixed in length, but with multiple unspecified codons can be cloned into genes III, VI, or VIII of bacteriophage M13 where they are expressed as a plurality of peptide:capsid fusion proteins. The libraries, often referred to as random peptide libraries, can be screened for binding to target molecules of interest. Usually, three to four rounds of screening can be accomplished in a week's time, leading to the isolation of one to hundreds of binding phage.

The primary structure of the binding peptides is then deduced by nucleotide sequencing of individual clones. Inspection of the peptide sequences sometimes reveals a common motif, or consensus sequence. Generally, this motif when synthesized as a soluble peptide has the full binding activity. Random peptide libraries have successfully yielded peptides that bind to the Fab site of antibodies (Cwirla et al., 1990, Proc. Natl. Acad. Sci. USA 87:6378-6382; Scott and Smith, 1990, Science 249:386-390), cell surface receptors (Doorbar and Winter, 1994, J. Mol. Biol. 244:361-369; Goodson et al., 1994, Proc. Natl. Acad. Sci. USA 91:7129-7133), cytosolic receptors (Blond-Elguindi et al., 1993, Cell 75:717-

728), intracellular proteins (Daniels and Lane, 1994, J. Mol. Biol. 243:639-652; Dedman et al., 1993, J. Biol. Chem. 268:23025-23030; Sparks et al., 1994, J. Biol. Chem. 269:23853-23856), DNA (Krook et al., 1994, Biochem. Biophys. Res. Comm. 204:849-854), and many other targets (Winter, 1994, Drug Dev. Res. 33:71-89).

Most vital cellular processes are regulated by the transmission of signals throughout the cell in the form of complex interactions between proteins. As the study of signal transduction, or the flow of information throughout the cell, has broadened and matured, it has become apparent that these protein-protein interactions are often mediated by modular domains within signalling proteins. Src, both the first proto-oncogene product and the first tyrosine kinase discovered (Taylor and Shalloway, 1993, Current Opinion in Genetics and Development 3:26-34), is the prototypic modular domain-containing protein.

Src is a protein tyrosine kinase of 60 kilodaltons and is located at the plasma membrane of cells. It was first discovered in the 1970's to be the oncogenic element of Rous sarcoma virus, and in the 1980's, it was appreciated to be a component of the signal transduction system in animal cells. However, since the identification of viral and cellular forms of Src (*i.e.*, v-Src and c-Src) , their respective roles in oncogenesis, normal cell growth, and differentiation have not been completely understood.

In addition to its tyrosine kinase region (sometimes called a Src Homology 1 domain), Src contains two regions that have been found to have functionally and structurally homologous counterparts in a large number of proteins. These regions have been designated the Src Homology 2 (SH2) and Src Homology 3 (SH3) domains. SH2 and SH3 domains are modular in that they fold independently of the protein that contains them, their secondary structure places N-and C-termini close to one another in space, and they appear at variable locations (anywhere from N-to C-terminal) from one protein to the next (Cohen et al., 1995, Cell 80:237-248). SH2 domains have been

well-studied and are known to be involved in binding to phosphorylated tyrosine residues (Pawson and Gish, 1992, Cell 71:359-362).

The Src-homology region 3 (SH3) of Src is a domain that is 60-70 amino acids in length and is present in many cellular proteins (Cohen et al., 1995, Cell 80:237-248; Pawson, 1995, Nature 373:573-580). Within Src, the SH3 domain is considered to be a negative inhibitory domain, because c-Src can be activated (*i.e.*, transforming) through mutations in this domain (Jackson et al., 1993, Oncogene 8:1943-1956; Seidel-Dugan et al., 1992, Mol Cell Biol 12:1835-1845).

To deduce the binding specificity of the Abl SH3 domain, a group led by David Baltimore screened cDNA libraries with radiolabeled GST-Abl SH3 fusion protein and identified two binding cDNA clones (Cicchetti et al., 1992, Science 257:803-806). Both clones encoded proteins with proline rich regions that were later shown to be SH3 binding domains.

Subsequently, others have screened combinatorial peptide libraries and identified peptides that bound to the Src SH3 domain (Yu et al., 1994, Cell 76:933-945; Cheadle et al., 1994, J. Biol. Chem. 269:24034-24039). Using the SH3 domain of Src, Sparks et al., 1994, J. Biol. Chem. 269:23853-23856 screened phage-display random peptide libraries and identified a consensus peptide sequence that binds with specificity and high affinity to the Src SH3 domain.

The consensus from these various studies is that the optimal Src SH3 peptide ligand is RPLPPLP (SEQ ID NO:45). Recently, the structures of the peptide-SH3 domain complexes have been deduced by NMR and the peptides have been shown to bind in two possible orientations with respect to the SH3 domain (Feng et al., 1994, Science 266:1241-1247; Lim et al., 1994, Nature 372:375-379).

Since SH3 domains have been found to have such important roles in the function of crucial signalling and structural elements in the cell, a method of identifying proteins containing SH3 regions is of great interest. In this regard, it is important to note that such a method is

unavailable because of the low sequence similarity of modular functional domains, including SH3. See, *e.g.*, Figure 6, which illustrates the minimal primary sequence homology among various known SH3 domains.

Sequence homology searches can potentially identify known proteins containing not yet recognized functional domains of interest, however, sequence homology generally needs to be >40% for this procedure to be successful. Functional domains generally are less than 40% homologous and therefore many would be missed in a sequence homology search. In addition, homology searches do not identify novel proteins; they only identify proteins already defined by nucleotide or amino acid sequence and present in the database.

Another approach is to use hybridization techniques using nucleotide probes to search expression libraries for novel proteins. This method would have limited applicability to finding novel proteins containing functional domains due to the low sequence homology of the functional domains.

Methods for isolating partner proteins involved in protein-protein interactions have generally focused on finding a ligand to a protein that has been found and characterized. Such approaches have included using anti-idiotypic antibodies that mimic the known protein to screen cDNA expression libraries for a binding ligand (Jerne, 1974, Ann. Immunol. (Inst. Pasteur) 125c:373-389; Sudol, 1994, Oncogene 9:2145-2152). Skolnick et al., 1991, Cell 65:83-90 isolated a binding partner for PI3-kinase by screening a cDNA expression library with the ^{32}P-labeled tyrosine phosphorylated carboxyl terminus of the epidermal growth factor receptor (EGFR).

An easy method for isolating operationally defined ligands involved in protein-protein interactions and for optimally identifying an exhaustive set of modular domain-containing proteins implicated in binding with the ligands would be highly desirable.

If such a method were available, however, such a method would be useful for the isolation of any polypeptide having a functioning version of *any* functional domain of

interest. Such a general method would be of tremendous utility in that whole families of related proteins each with its own version of the functional domain of interest could be identified. Knowledge of such related proteins would contribute greatly to our understanding of various physiological processes, including cell growth or death, malignancy, and immune reactions, to name a few. Such a method would also contribute to the development of increasingly more effective therapeutic, diagnostic, or prophylactic agents having fewer side effects.

According to the present invention, just such a method is provided.

Regarding SH3 domain-containing proteins, the method of the present invention will contribute greatly to our understanding of cell growth (Zhu et al., 1993, J. Biol. Chem. 268:1775-1779; Taylor and Shalloway, 1994, Nature 368:867-871), malignancy (Wages et al., 1992, J. Virol. 66:1866-1874; Bruton and Workman, 1993, Cancer Chemother. Pharmacol. 32:1-19), subcellular localization of proteins to the cytoskeleton and/or cellular membranes (Weng et al., 1993, J. Biol. Chem. 268:14956-14963; Bar-Sagi et al., 1993, Cell 74:83-91), signal transduction (Duchesne et al., 1993, Science 259:525-528), cell morphology (Wages et al., 1992, J. Virol. 66:1866-1874; McGlade et al., 1993, EMBO J. 12:3073-3081), neuronal differentiation Tanaka et al., 1993, Mol. Cell. Biol. 13:4409-4415), T cell activation (Reynolds et al., 1992, Oncogene 7:1949-1955), and cellular oxidase activity (McAdara and Babior, 1993, Blood 82:A28).

Citation of a reference hereinabove shall not be construed as an admission that such is prior art to the present invention.

3. SUMMARY OF THE INVENTION

In general, the present invention is directed to a method of using isolated, operationally defined ligands involved in binding interactions for optimally identifying an

exhaustive set of compounds binding to such ligands. In one embodiment, the isolated ligands are peptides involved in specific protein-protein interactions and are used to identify a set of novel modular domain-containing proteins that bind to the ligands. Using this method, proteins sharing only modest similarities but a common function can be found.

The present invention is directed to a method of identifying a polypeptide or family of polypeptides having a functional domain of interest. The basic steps of the method comprise: (a) choosing a recognition unit or set of recognition units having a selective affinity for a target molecule with a functional domain of interest; (b) contacting the recognition unit with a plurality of polypeptides; and (c) identifying a polypeptide having a selective binding affinity for the recognition unit, which polypeptide includes the functional domain of interest or a functional equivalent thereof.

In one particular embodiment of the invention, exhaustive screening of proteins having a desired functional domain involves an iterative process by which ligands or recognition units for SH3 domains identified in the first round of screening are used to detect SH3 domain-containing proteins in successive expression library screens.

More particularly, the method of the present invention includes choosing a recognition unit having a selective affinity for a target molecule with a functional domain of interest. With this recognition unit (particularly under the multvalent recognition unit screening conditions taught by the present invention), it has further been discovered that a plurality of polypeptides from various sources can be examined such that certain polypeptides having a selective binding affinity for the recognition unit can be identified. The polypeptides so identified have been shown to include the functional domain of interest; that is, the functional domains found are working versions that are capable of displaying the same binding specificity as the functional domain of interest. Hence, the polypeptides identified by the

present method also possess those attributes of the functional domain of interest which allow these related polypeptides to exhibit the same, similar, or analogous (but functionally equivalent) selective affinity characteristics as the domain of interest of the initial target molecule. By screening the plurality of peptides for recognition unit binding, the methods of the present invention circumvent the limitations of conventional DNA-based screening methods and allow for the identification of highly disparate protein sequences possessing functionally equivalent functional domains.

In specific embodiments of the present invention, the plurality of polypeptides is obtained from the proteins present in a cDNA expression library. The specificity of the polypeptides which bear the functional domain of interest or a functional equivalent thereof for various peptides or recognition units can subsequently be examined, allowing for a greater understanding of the physiological role of particular polypeptide/recognition unit interactions. Indeed, the present invention provides a method of targeted drug discovery based on the observed effects of a given drug candidate on the interaction between a recognition unit-polypeptide pair or a recognition unit and a "panel" of related polypeptides each with a copy or a functional equivalent of (*e.g.*, capable of displaying the same binding specificity and thus binding to the same recognition unit as) the functional domain of interest.

The present invention also provides polypeptides comprising certain amino acid sequences. Moreover, the present invention also provides nucleic acids, including certain DNA constructs comprising certain coding sequences. Using the methods of the present invention, more than eighteen different SH3 domain-containing proteins have been identified, over half of which have not been previously described.

The present inventors have found, unexpectedly, that the valency (*i.e.*, whether it is a monomer, dimer, tetramer, *etc.*) of the recognition unit that is used to screen an expression library or other source of polypeptides apparently

has a marked effect upon the specificity of the recognition unit-functional domain interaction. The present inventors have discovered that recognition units in the form of small peptides, in multivalent form, have a specificity that is eased but not forfeited. In particular, biotinylated peptides bound to a multivalent (believed to be tetravalent) streptavidin-alkaline phosphatase complex have an unexpected generic specificity. This allows such peptides to be used to screen libraries to identify classes of polypeptides containing functional domains that are similar but not identical in sequence to the peptides' original target functional domains.

The present invention also provides methods for identifying potential new drug candidates (and potential lead compounds) and determining the specificities thereof. For example, knowing that a polypeptide with a functional domain of interest and a recognition unit, *e.g.*, a binding peptide, exhibit a selective affinity for each other, one may attempt to identify a drug that can exert an effect on the polypeptide-recognition unit interaction, *e.g.*, either as an agonist or as an antagonist (inhibitor) of the interaction. With this assay, then, one can screen a collection of candidate "drugs" for the one exhibiting the most desired characteristic, *e.g.*, the most efficacious in disrupting the interaction or in competing with the recognition unit for binding to the polypeptide.

In addition, the present invention also provides certain assay kits and methods of using these assay kits for screening drug candidates for their ability to affect the binding of a polypeptide containing a functional domain to a recognition unit. In a particular aspect of the present invention, the assay kit comprises: (a) a polypeptide containing a functional domain of interest; and (b) a recognition unit having a selective binding affinity for the polypeptide. Yet another assay kit may comprise a plurality of polypeptides, each polypeptide containing a functional domain of interest, in which the functional domain of interest

is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix, and at least one recognition unit having a selective affinity for each of the plurality of polypeptides.

Other objects of the present invention will be apparent to those of ordinary skill upon further consideration of the following detailed description.

4. DESCRIPTION OF THE FIGURES

Figure 1 is a schematic representation of the general aspects of a method of identifying recognition units exhibiting a selective affinity for a target molecule with a functional domain of interest. In this illustration, the target molecule is a polypeptide with an SH3 domain, and the recognition units are peptides having a selective affinity for the SH3 domain that are expressed in a phage displayed library.

Figure 2 illustrates the selectivities exhibited by particular recognition units that bind to the Src SH3 domain (in this case, two heptapeptides) for a "panel" of known polypeptides known to contain an SH3 domain. The non-SH3-containing protein, GST, serves as control. RPLPPLP is (SEQ ID NO:45); APPVPPR is (SEQ ID NO:203)

Figure 3 is a schematic representation of the general method of identifying polypeptides with a functional domain of interest by screening a plurality of polypeptides using a suitable recognition unit. In the illustration, the plurality of polypeptides is obtained from a cDNA expression library, and the recognition units are SH3 domain-binding peptides.

Figure 4 illustrates how an SH3 domain-binding peptide can be used to identify other SH3 domain-containing proteins. Shown is a schematic representation of the

progression from initial selection of a target molecule with a functional domain of interest, choice of recognition unit, and identification of polypeptides that have a selective affinity for the recognition unit and include the functional domain of interest or a functional equivalent thereof.

Figure 5 depicts filters from primary (Figure 5B) and tertiary (Figure 5A) screens of a λcDNA library probed with a biotinylated SH3-binding peptide recognition unit in the form of a complex with streptavidin-alkaline phosphatase (SA-AP). A mouse 16 day embryo cDNA library in λEX*lox* was incubated with a multivalent complex formed between biotinylated pSrcCII and SA-AP. The sites of peptide binding were detected by incubation with BCIP (5-bromo-4-chloro-3-indoyl-phosphate-p-toluidine salt) and NBT (nitroblue tetrazolium chloride) for approximately five minutes.

Figure 6 shows an alignment of SH3 domains that illustrates the minimal primary sequence homology among various known SH3 domains. The amino acid sequences shown are SEQ ID NOs:68-111.

Figure 7A is a schematic representation of a population of functional domains represented by the circles. "A" is a recognition unit specific to one circle only. B, on the other hand, recognizes three domains, while B1 and B2 recognize only two each. Figure 7B illustrates an iterative method whereby new recognition units are chosen based on polypeptides uncovered with the first recognition unit(s). These new recognition units lead to the identification of other related polypeptides, *etc.*, expanding the scope of the study to increasingly diverse members of the related population.

Figure 8 illustrates the binding specificity of several SH3 domain recognition units. Biotinylated Class I (pSrcCI) or Class II (pSrcCII) Src SH3 domain recognition

units, Crk SH3 domain recognition units (pCrk), PLCγ SH3 domain recognition units (pPLC), and Abl SH3 domain recognition units (pAbl) were tested for binding to the indicated GST-SH3 domain fusion proteins immobilized onto duplicate microtiter plate wells. Recognition units are listed along the left side of the figure; GST-SH3 domain fusion proteins are listed along the bottom. Recognition units were incubated either as multivalent complexes of biotinylated peptides and streptavidin-horseradish peroxidase (SA-HRP) (complexed) or as monovalent biotinylated peptides (uncomplexed), followed by incubation with SA-HRP. Average optical densities are shown.

Figure 9 shows a schematic of SH3-domain containing proteins isolated using the present invention. The name, identity, type of screen, and number of individual clones derived for each sequence are indicated. Diagrams are to scale, with SH3 domains representing approximately 60 amino acids. The abbreviations AR, P, CR, E/P, and SH2 represent ankyrin repeats, proline-rich segments, Cortactin repeats, glutamate/proline-rich segments, and Src homology 2 domains, respectively. Flared ends represent putative translation initiation sites for individual cDNAs. The Mouse, Human 1, and Human 2 libraries correspond to mouse 16 day embryo, human bone marrow, and human prostate cancer cDNA libraries, respectively. For a description of the pSrcII and pCort recognition units, see Section 6.1.

Figure 10A and 10B depicts the sequence alignment of SH3 domains in proteins isolated using the present invention. The name and identity of each clone is indicated. Where appropriate, multiple SH3 domains from the same polypeptide are designated A, B, C, etc., from N- to C-terminal. Periods indicate gaps introduced to maximize alignment of similar residues. Positions corresponding to conserved residues shown to be involved in ligand binding in the SH3 domains of Src and Grb2/Sem5 (Tomasetto et al., 1995, Genomics 28:367-376) are

presented in bold and underlined, respectively. Primary structures of SH3P1-8 and SH3P10-13 correspond to mouse, SH3P15-18, clone 5, 34, 40, 41, 45, 53, 55, 56, and 65 to human, and SH3P9 and SH3P14 to mouse (m) or human (h) cDNA clones. For sequence comparison, the sequence of the mouse c-Src SH3 domain (GenBank accession number P41240) is shown. The GenBank accession numbers for mouse Cortactin, SPY75/HS1, Crk, and human MLN50, Lyn, Fyn, and Src are U03184, D42120, S72408, X82456, M16038, P06241, and P41240, respectively. The amino acid sequences shown are SEQ ID NOs:112-140.

Figure 11 depicts the specificity continuum described in Section 5.2.1. "SA-AP peptide complex" represents the multivalent (believed to be tetravalent) complex of streptavidin-alkaline phosphatase and biotinylated peptide described in that section.

Figure 12 depicts the results of experiments in which peptide recognition units were synthesized and tested for their ability to bind to novel SH3 domains described in Sections 6.1 and 6.1.1. A minus indicates no binding; a plus indicates binding, with the number of pluses indicating the strength of binding. For further details, see Section 6.2. The amino acid sequences shown are SEQ ID NOs:141-168.

Figure 13 depicts more data from the experiment depicted in Figure 12. The amino acid sequences shown are SEQ ID NOs:169-188.

Figure 14 illustrates the effect of preconjugation with streptavidin-alkaline phosphatase on the affinity of biotinylated peptides for SH3 domains. See Section 6.3.1 for details.

Figure 15 illustrates the effect of preconjugation with streptavidin-alkaline phosphatase on the specificity of biotinylated peptides for GST-SH3 domain fusion proteins that

have been immobilized on nylon membranes. See Section 6.3.2 for details.

Figure 16 illustrates the effect of preconjugation with streptavidin-alkaline phosphatase on the specificity of biotinylated peptides for proteins containing SH3 domains expressed by cDNA clones. See Section 6.3.3 for details.

Figure 17 illustrates a strategy for exhaustively screening an expression library for SH3 domain-containing proteins. A peptide recognition unit is generated by screening a combinatorial peptide library for binders to an SH3 domain espressed bacterially as a GST fusion protein. This peptide is then used as a multivalent streptavidin-biotinylated peptide complex to screen for a subset of the SH3 domain-containing proteins represented in a cDNA expression library. A combinatorial library is once again used to identify recognition units of SH3 domains identified in the first expression library screen; these recognition units identify overlapping sets of proteins from the expression library. With multiple iterations of this process, it should be possible to clone systematically all SH3 domains represented in a given cDNA expression library.

Figure 18 depicts the nucleotide sequence of SH3P1, mouse p53bp2 (SEQ ID NO:5).

Figure 19 depicts the amino acid sequence of SH3P1, mouse p53bp2 (SEQ ID NO:6).

Figure 20 depicts the nucleotide sequence of SH3P2, a novel mouse gene (SEQ ID NO:7).

Figure 21 depicts the amino acid sequence of SH3P2, a novel mouse gene (SEQ ID NO:8).

Figure 22 depicts the nucleotide sequence of SH3P3, a novel mouse gene (SEQ ID NO:9).

Figure 23 depicts the amino acid sequence of SH3P3, a novel mouse gene (SEQ ID NO:10).

Figure 24 depicts the nucleotide sequence of SH3P4, a novel mouse gene (SEQ ID NO:11).

Figure 25 depicts the amino acid sequence of SH3P4, a novel mouse gene (SEQ ID NO:12).

Figure 26 depicts the nucleotide sequence of SH3P5, mouse Cortactin (SEQ ID NO:13).

Figure 27 depicts the amino acid sequence of SH3P5, mouse Cortactin (SEQ ID NO:14).

Figure 28 depicts the nucleotide sequence of SH3P6, mouse MLN50 (SEQ ID NO:15).

Figure 29 depicts the amino acid sequence of SH3P6, mouse MLN50 (SEQ ID NO:16).

Figure 30 depicts the nucleotide sequence of SH3P7, a novel mouse gene (SEQ ID NO:17).

Figure 31 depicts the amino acid sequence of SH3P7, a novel mouse gene (SEQ ID NO:18).

Figure 32 depicts the nucleotide sequence of SH3P8, a novel mouse gene (SEQ ID NO:19).

Figure 33 depicts the amino acid sequence of SH3P8, a novel mouse gene (SEQ ID NO:20).

Figure 34 depicts the nucleotide sequence of SH3P9, a novel mouse gene (SEQ ID NO:21).

Figure 35 depicts the amino acid sequence of SH3P9, a novel mouse gene (SEQ ID NO:22).

Figure 36 depicts the nucleotide sequence of SH3P9, a novel human gene (SEQ ID NO:23).

Figure 37 depicts the amino acid sequence of SH3P9, a novel human gene (SEQ ID NO:24).

Figure 38 depicts the nucleotide sequence of SH3P10, mouse HS1 (SEQ ID NO:25).

Figure 39 depicts the amino acid sequence of SH3P10, mouse HS1 (SEQ ID NO:26).

Figure 40 depicts the nucleotide sequence of SH3P11, mouse Crk (SEQ ID NO:27).

Figure 41 depicts the amino acid sequence of SH3P11, mouse Crk (SEQ ID NO:28).

Figure 42A depicts the nucleotide sequence from positions 1-2600 of SH3P12, a novel mouse gene (a portion of SEQ ID NO:29).

Figure 42B depicts the nucleotide sequence from positions 2601-3335 of SH3P12, a novel mouse gene (a portion of SEQ ID NO:29).

Figure 43 depicts the amino acid sequence of SH3P12, a novel mouse gene (SEQ ID NO:30).

Figure 44 depicts the nucleotide sequence of SH3P13, a novel mouse gene (SEQ ID NO:31).

Figure 45 depicts the amino acid sequence of SH3P13, a novel mouse gene (SEQ ID NO:32).

Figure 46A depicts the nucleotide sequence from positions 1-2400 of SH3P14, mouse H74 (a portion of SEQ ID NO:33).

Figure 46B depicts the nucleotide sequence from positions 2351-4091 of SH3P14, mouse H74 (a portion of SEQ ID NO:33).

Figure 47 depicts the amino acid sequence of SH3P14, mouse H74 (SEQ ID NO:34).

Figure 48 depicts the nucleotide sequence of SH3P14, human H74 (SEQ ID NO:35).

Figure 49 depicts the amino acid sequence of SH3P14, human H74 (SEQ ID NO:36).

Figure 50 depicts the nucleotide sequence of SH3P17, a novel human gene (SEQ ID NO:37).

Figure 51 depicts the amino acid sequence of SH3P17, a novel human gene (SEQ ID NO:38).

Figure 52A depicts the nucleotide sequence of SH3P18, a novel human gene (SEQ ID NO:39).

Figure 53 depicts the amino acid sequence of SH3P18, a novel human gene (SEQ ID NO:40).

Figure 54 depicts the nucleotide sequence of clone 55, a novel human gene (SEQ ID NO:189).

Figure 55 depicts the amino acid sequence of clone 55, a novel human gene (SEQ ID NO:190).

Figure 56 depicts the nucleotide sequence of clone 56, a novel human gene (SEQ ID NO:191).

Figure 57 depicts the amino acid sequence of clone 56, a novel human gene (SEQ ID NO:192).

Figure 58A depicts the nucleotide sequence from position 1-1720 of clone 65, a novel human gene (a portion of SEQ ID NO:193).

Figure 58B depicts the nucleotide sequence from position 1721-2873 of clone 65, a novel human gene (a portion of SEQ ID NO:193).

Figure 59 depicts the amino acid sequence of clone 65, a novel human gene (SEQ ID NO:194).

Figure 60 depicts the nucleotide sequence of clone 34, a novel human gene (SEQ ID NO:195).

Figure 61A depicts a portion of the amino acid sequence of clone 34, a novel human gene (a portion of SEQ ID NO:196).

Figure 61B depicts a portion of the amino acid sequence of clone 34, a novel human gene (a portion of SEQ ID NO:196).

Figure 62 depicts the nucleotide sequence of clone 41, a novel human gene (SEQ ID NO:197).

Figure 63A depicts a portion of the amino acid sequence of clone 41, a novel human gene (a portion of SEQ ID NO:198).

Figure 63B depicts a portion of the amino acid sequence of clone 41, a novel human gene (a portion of SEQ ID NO:198).

Figure 64A depicts the nucleotide sequence of clone 53, a novel human gene (SEQ ID NO:199).

Figure 65A depicts a portion of the amino acid sequence of clone 53, a novel human gene (a portion of SEQ ID NO:200).

Figure 65B depicts a portion of the amino acid sequence of clone 53, a novel human gene (a portion of SEQ ID NO:200).

Figure 66A and 66B depicts the nucleotide sequence (SEQ ID NO:220) and amino acid sequence (SEQ ID NO:221) of clone 5, a novel human gene.

5. DETAILED DESCRIPTION OF THE INVENTION

As stated above, the present invention is related broadly to certain polypeptides having a functional domain of interest and is directed to methods of identifying and using these polypeptides. The present invention is also directed to a method of using isolated, operationally defined ligands involved in binding interactions for optimally identifying an exhaustive set of compounds binding such ligands and to compounds, target molecules, and, in one embodiment, polypeptides having a functional domain of interest and to methods of using these compounds. The detailed description that follows is provided to elucidate the invention further and to assist further those of ordinary skill who may be interested in practicing particular aspects of the invention.

First, certain definitions are in order. Accordingly, the term "polypeptide" refers to a molecule comprised of amino acid residues joined by peptide (*i.e.*, amide) bonds and includes proteins and peptides. Hence, the

polypeptides of the present invention may have single or multiple chains of covalently linked amino acids and may further contain intrachain or interchain linkages comprised of disulfide bonds. Some polypeptides may also form a subunit of a multiunit macromolecular complex. Naturally, the polypeptides can be expected to possess conformational preferences and to exhibit a three-dimensional structure. Both the conformational preferences and the three-dimensional structure will usually be defined by the polypeptide's primary (*i.e.*, amino acid) sequence and/or the presence (or absence) of disulfide bonds or other covalent or non-covalent intrachain or interchain interactions.

The polypeptides of the present invention can be any size. As can be expected, the polypeptides can exhibit a wide variety of molecular weights, some exceeding 150 to 200 kilodaltons (kD). Typically, the polypeptides may have a molecular weight ranging from about 5,000 to about 100,000 daltons. Still others may fall in a narrower range, for example, about 10,000 to about 75,000 daltons, or about 20,000 to about 50,000 daltons.

The phrase "functional domain" refers to a region of a polypeptide which affords the capacity to perform a particular function of interest. This function may give rise to a biological, chemical, or physiological consequence that may be reversible or irreversible and which may include, but not be limited to, protein-protein interactions (*e.g.*, binding interactions) involving the functional domain, a change in the conformation or a transformation into a different chemical state of the functional domain or of molecules acted upon by the functional domain, the transduction of an intracellular or intercellular signal, the regulation of gene or protein expression, the regulation of cell growth or death, or the activation or inhibition of an immune response. Furthermore, the functional domain *of interest* is defined by a particular functional domain that is present in a given target molecule. A discussion of the selection of a particular functional domain-containing target molecule is presented further below.

Many functional domains tend to be modular in that such domains may occur one or more times in a given polypeptide (or target molecule) or may be found in a family of different polypeptides. When found more than once in a given polypeptide or in different polypeptides, the modular functional domain may possess substantially the same structure, in terms of primary sequence and/or three-dimensional space, or may contain slight or great variations or modifications among the different versions of the functional domain of interest.

What is important, however, is that these related functional domains retain the functional aspects of the functional domain of interest present in the target molecule. It is stressed that, indeed, it is this functional relationship among two or more possible versions of a functional domain of interest which may be identified, defined, and exploited by the methods of the present invention. In a preferred aspect, the function of interest is the ability to bind to a molecule (*e.g.*, a peptide) of interest.

The present invention provides a general strategy by which recognition units that bind to a functional domain-containing molecule can be used to screen expression libraries of genes (*e.g.*, cDNA, genomic libraries) systematically for novel functional domain-containing proteins. In specific embodiments, the recognition units are prior isolated from a random peptide library, or are known peptide ligands or recognition units, or are recognition units that are identified by database searches for sequences having homology to a peptide recognition unit having the binding specificity of interest. Using the methods of the present invention, it is possible to exhaustively screen an expression library for proteins with a given functional domain.

In the prior art, novel genes (and thus their encoded protein products) are most commonly identified from cDNA libraries. Generally, an appropriate cDNA library is screened with a probe that is either an oligonucleotide or an

antibody. In either case, the probe must be specific enough for the gene that is to be identified to pick that gene out from a vast background of non-relevant genes in the library. It is this need for a specific probe that is the highest hurdle that must be overcome in the prior art identification of novel genes. Another method of identifying genes from cDNA libraries is through use of the polymerase chain reaction (PCR) to amplify a segment of a desired gene from the library. PCR requires that oligonucleotides having sequence similarity to the desired gene be available.

If the probe used in prior art methods is a nucleic acid, the cDNA library may be screened without the need for expressing any protein products that might be encoded by the cDNA clones. If the probe used in prior art methods is an antibody, then it is necessary to build the cDNA library into a suitable expression vector. For a comprehensive discussion of the art of identifying genes from cDNA libraries, see Sambrook, Fritsch, and Maniatis, "Construction and Analysis of cDNA Libraries," Chapter 8 in Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, 1989. See also Sambrook, Fritsch, and Maniatis, "Screening Expression Libraries with Antibodies and Oligonucleotides," Chapter 12 in Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, 1989.

As an alternative to cDNA libraries, genomic libraries are used. When genomic libraries are used in prior art methods, the probe is virtually always a nucleic acid probe. See Sambrook, Fritsch, and Maniatis, "Analysis and Cloning of Eukaryotic Genomic DNA," Chapter 9 in Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, 1989.

In the prior art, nucleic acid probes used in screening libraries are often based upon the sequence of a known gene that is thought to be homologous to a gene that it is desired to isolate. The success of the procedure depends upon the degree of homology between the probe and the target gene being sufficiently high. Probes based upon the sequences

of known functional domains in proteins had limited value because, while the sequences of the functional domains were similar enough to allow for their recognition as shared domains, the similarity was not so high that the probes could be used to screen cDNA or genomic libraries for genes containing the functional domains.

PCR may also be used to identify genes from genomic libraries. However, as in the case of using PCR to identify genes from cDNA libraries, this requires that oligonucleotides having sequence similarity to the desired gene be available.

Using the screening methods provided by the present invention, DNA encoding proteins having a desired functional domain that would not be readily identified by sequence homology can be identified by functional binding specificity to recognition units. By virtue of an ease in specificity of binding requirements conferred by the screening methods of the present invention, many novel, functionally homologous, functional domain-containing proteins can be identified. Although not intending to be bound by any mechanistic explanation, this ease in binding specificity is believed to be the result of the use of a multivalent peptide recognition unit used to screen the gene library, preferably of a valency greater than bivalent, more preferably tetravalent or greater, and most preferably the streptavidin-biotinylated peptide recognition unit complex.

In one particular embodiment of the invention, exhaustive screening of proteins having a desired functional domain involves an iterative process by which recognition units for SH3 domains identified in the first round of screening are used to detect SH3 domain-containing proteins in successive expression library screens (see Figure 17). This strategy enables one to search "sequence space" in what might be thought of as ever-widening circles with each successive cycle. This iterative strategy can be initiated even when only one functional domain-containing protein and recognition unit are available.

This iterative process is not limited to proteins containing SH3 domains. Members within a class of other functional domains also tend to have overlapping, or at least similar recognition unit preferences, are structurally stable, and often confer similar binding properties to a wide variety of proteins. These characteristics predict that the methods of the present invention will be applicable to a wide variety of functional domain-containing proteins in addition to their applicability to SH3 domain-containing proteins.

5.1. Discovery of Novel Genes and Polypeptides Containing Functional Domains

The present invention provides methods for the identification of one or more polypeptides (in particular, a "family" of polypeptides, including the target molecule) that contains a functional domain of interest that either corresponds to or is the functional equivalent of a functional domain of interest present in a predetermined target molecule.

The present invention provides a mechanism for the rapid identification of genes (*e.g.*, cDNAs) encoding virtually any functional domain of interest. By screening cDNA libraries or other sources of polypeptides for recognition unit binding rather than sequence similarity, the present invention circumvents the limitations of conventional DNA-based screening methods and allows for the identification of highly disparate protein sequences possessing equivalent functional activities. The ability to isolate entire repertoires of proteins containing particular modular functional domains will prove invaluable both in molecular biological investigations of the genome and in bringing new targets into drug discovery programs.

It should likewise be apparent that a wide range of polypeptides having a functional domain of interest can be identified by the process of the invention, which process comprises:

(a) contacting a multivalent recognition unit complex with a plurality of polypeptides; and

(b) identifying a polypeptide having a selective binding affinity for said recognition unit complex.

In a specific embodiment, the process comprises:

(a) contacting a multivalent recognition unit complex with a plurality of polypeptides from which it is desired to identify a polypeptide having selective binding affinity for the recognition unit, in which the valency of the recognition unit in the complex is at least two, or at least four; and

(b) identifying, and preferably recovering, a polypeptide having a selective binding affinity for the recognition unit complex.

In another specific embodiment, the process comprises a method of identifying at least one polypeptide comprising a functional domain of interest, said method comprising:

(a) contacting one or more multivalent recognition unit complexes with a plurality of polypeptides; and

(b) identifying at least one polypeptide having selective binding affinity for at least one of said recognition unit complexes.

In another specific embodiment, the process comprises:

(a) contacting a multivalent recognition unit complex, which complex comprises (i) avidin or streptavidin, and (ii) biotinylated recognition units, with a plurality of polypeptides from a cDNA expression library, in which the recognition unit is a peptide having in the range of 6 to 60 amino acid residues; and

(b) identifying a polypeptide having a selective binding affinity for said recognition unit complex.

In another specific embodiment, the process comprises a method of identifying a polypeptide having an SH3 domain of interest comprising:

(a) contacting a multivalent recognition unit complex, which complex comprises (i) avidin or streptavidin, and (ii) biotinylated recognition units, with a plurality of

polypeptides from a cDNA expression library, in which the recognition unit is a peptide having in the range of 6 to 60 amino acid residues and which selectively binds an SH3 domain; and

(b) identifying a polypeptide having a selective binding affinity for said recognition unit complex.

In another specific embodiment, the process comprises a method of identifying a polypeptide having a functional domain of interest or a functional equivalent thereof comprising:

(a) screening a random peptide library to identify a peptide that selectively binds a functional domain of interest; and

(b) screening a cDNA or genomic expression library with said peptide or a binding portion thereof to identify a polypeptide that selectively binds said peptide.

In a specific embodiment of the above method, the screening step (b) is carried out by use of said peptide in the form of multiple antigen peptides (MAP) or by use of said peptide cross-linked to bovine serum albumin or keyhole limpet hemocyanin.

In another specific embodiment, the process comprises a method of identifying a polypeptide having a functional domain of interest or a functional equivalent thereof comprising:

(a) screening a random peptide library to identify a plurality of peptides that selectively bind a functional domain of interest;

(b) determining at least part of the amino acid sequences of said peptides;

(c) determining a consensus sequence based upon the determined amino acid sequences of said peptides; and

(d) screening a cDNA or genomic expression library with a peptide comprising the consensus sequence to identify a polypeptide that selectively binds said peptide.

In another specific embodiment, the process comprises a method of identifying a polypeptide having a

functional domain of interest or a functional equivalent thereof comprising:

(a) screening a random peptide library to identify a first peptide that selectively binds a functional domain of interest;

(b) determining at least part of the amino acid sequence of said first peptide;

(c) searching a database containing the amino acid sequences of a plurality of expressed natural proteins to identify a protein containing an amino acid sequence homologous to the amino acid sequence of said first peptide; and

(d) screening a cDNA or genomic expression library with a second peptide comprising the sequence of said protein that is homologous to the amino acid sequence of said first peptide.

The identified polypeptide identified by the above-described methods thus should contain the functional domain of interest or a functional equivalent thereof (that is, having a functional domain that is identical, or having a functional domain that differs in sequence but is capable of binding to the same recognition unit). In a particular embodiment, the polypeptide identified is a novel polypeptide. In a preferred embodiment, the recognition unit that is used to form the multvalent recognition unit complex is isolated or identified from a random peptide library.

In a specific embodiment, the present invention provides amino acid sequences and DNA sequences encoding novel proteins containing SH3 domains. The SH3 domains vary in sequence but retain binding specificity to an SH3 domain recognition unit. Also provided are fragments and derivatives of the novel proteins containing SH3 domains as well as DNA sequences encoding the same. It will be apparent to one of ordinary skill in the art that also provided are proteins that vary slightly in sequence from the novel proteins by virtue of conservative amino acid substitutions. It will also be apparent to one of ordinary skill in the art that the novel

proteins may be expressed recombinantly by standard methods. The novel proteins may also be expressed as fusion proteins with a variety of other proteins, *e.g.*, glutathione S-transferase.

The present invention provides a purified polypeptide comprising an SH3 domain, said SH3 domain having an amino acid sequence selected from the group consisting of: SEQ ID NOs: 113-115, 118-121, 125-128, 133-139, 204-218, and 219. Also provided is a purified DNA encoding the polypeptide.

Also provided is a purified polypeptide comprising an SH3 domain, said polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221. Also provided is a purified DNA encoding the polypeptide.

Also provided is a purified DNA encoding an SH3 domain, said DNA having a sequence selected from the group consisting of SEQ ID NOs: 7, 9, 11, 17, 19, 21, 23, 29, 31, 37, 39, 189, 191, 193, 195, 197, 199, and 220. Also provided is a nucleic acid vector comprising this purified DNA. Also provided is a recombinant cell containing this nucleic acid vector.

Also provided is a purified DNA encoding a polypeptide having an amino acid sequence selected from the group consisting of: SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221. Also provided is a nucleic acid vector comprising this purified DNA. Also provided is a recombinant cell containing this nucleic acid vector.

Also provided is a purified DNA encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: SEQ ID NOs:113-115, 118-121, 125-128, 133-139, 204-218, and 219. Also provided is a nucleic acid vector comprising this purified DNA. Also provided is a recombinant cell containing this nucleic acid vector.

Also provided is a purified molecule comprising an SH3 domain of a polypeptide having an amino acid sequence selected from the group consisting of: SEQ ID NO: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

Also provided is a fusion protein comprising (a) an amino acid sequence comprising an SH3 domain of a polypeptide having the amino acid sequence of SEQ ID NO: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221 joined via a peptide bond to (b) an amino acid sequence of at least six, or ten, or twenty amino acids from a different polypeptide. Also provided is a purified DNA encoding the fusion protein. Also provided is a nucleic acid vector comprising the purified DNA encoding the fusion protein. Also provided is a recombinant cell containing this nucleic acid vector. Also provided is a method of producing this fusion protein comprising culturing a recombinant cell containing a nucleic acid vector encoding said fusion protein such that said fusion protein is expressed, and recovering the expressed fusion protein.

The present invention also provides a purified nucleic acid hybridizable to a nucleic acid having a sequence selected from the group consisting of: SEQ ID NOs: 7, 9, 11, 17, 19, 21, 23, 29, 31, 37, 39, 189, 191, 193, 195, 197, 199, and 220.

The present invention also provides antibodies to a polypeptide having an amino acid sequence selected from the group consisting of: SEQ ID NOs:113-115, 118-121, 125-128, 133-139, 204-218, and 219.

The present invention also provides antibodies to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

It is demonstrated by way of example herein that recognition units that comprise SH3 domain ligands derived from combinatorial peptide libraries may be used in the methods of the present invention as probes for the rapid

discovery of novel proteins containing SH3 functional domains. The methods of the present invention require no prior knowledge of the characteristics of a SH3 domain's natural cellular ligand to initiate the process of discovery. One needs only enough purified SH3 domain-containing protein (by way of example, 1-5μg) to select peptides from a random peptide library. In addition, because the methods of the present invention identify novel proteins from cDNA expression libraries based only on their binding properties, low primary sequence identity between the target SH3 domain and the SH3 domains of the novel proteins discovered need not be a limitation, provided some functional similarity between these SH3 domains is conserved. Also, the methods of the present invention are rapid, require inexpensive reagents, and employ simple and well established laboratory techniques.

Using these methods, more than eighteen different SH3 domain-containing proteins have been identified, over half of which have not been previously described. While certain of these previously unknown proteins are clearly related to known genes such as amphiphysin and drebrin, others constitute new classes of signal transduction and/or cytoskeletal proteins. These include SH3P17 and SH3P18, two members of a new family of adaptor-like proteins comprised of multiple SH3 domains; SH3P12, a novel protein with three SH3 domains and a region similar to the extracellular peptide hormone sorbin; and SH3P4, SH3P8, and SH3P13, three members of a third new family of SH3-containing proteins. These novel proteins are described more fully in Sections 6.1 and 6.1.1. The high incidence of novel proteins identified by the methods of the present invention indicates that a large number of SH3 domain-containing proteins remain to be discovered by application of the methods of the invention.

One of ordinary skill in the art would recognize that the above-described novel proteins need not be used in their entirety in the various applications of those proteins described herein. In many cases it will be sufficient to employ that portion of the novel protein that contains the

functional (e.g., SH3) domain. Such exemplary portions of SH3 domain-containing proteins are shown in Figure 10A and 10B. Accordingly, the present invention provides derivatives (e.g., fragments and molecules comprising these fragments) of novel proteins that contain SH3 domains, e.g., as shown in Figure 10A and 10B. Nucleic acids encoding these fragments or other derivatives are also provided.

In another embodiment, the present invention includes a method of identifying one or more novel polypeptides having an SH3 domain, said method comprising:

(a) identifying a recognition unit having a selective affinity for the SH3 domain by screening a peptide library with the SH3 domain;

(b) producing said recognition unit;

(c) contacting said recognition unit with a source of polypeptides; and

(d) identifying one or more novel polypeptides having a selective affinity for said recognition unit, which polypeptides comprise the SH3 domain.

5.1.1 Functional Domains

Functional domains of interest in the practice of the present invention can take many forms and may perform a variety of functions. For example, such functional domains may be involved in a number of cellular, biochemical, or physiological processes, such as cellular signal transduction, transcriptional regulation, translational regulation, cell adhesion, migration or transport, cytokine secretion and other aspects of the immune response, and the like. In particular embodiments of the present invention, the functional domains of interest may consist of regions known as SH1, SH2, SH3, PH, PTB, LIM, armadillo, and Notch/ankyrin repeat. See, e.g., Pawson, 1995, Nature 373:573-580; Cohen et al., 1995, Cell 80:237-248. Functional domains may also be chosen from among regions known as zinc fingers, leucine zippers, and helix-turn-helix or helix-loop-helix. Certain functional domains may be binding domains, such as DNA-binding domains or actin-

binding domains. Still other functional domains may serve as sites of catalytic activity.

In one embodiment of the invention, a suitable target molecule containing the chosen functional domain of interest is selected. In the case of an SH3 domain, for example, a number of proteins (or functional domain-containing derivatives or analogs thereof) may be selected as the target molecule, including but not limited to, the Src family of proteins: Fyn, Lck, Lyn, Src, or Yes. Still other proteins contain an SH3 domain and can be used, including, but not limited to: Abl, Crk, Nck (other oncogenes), Grb2, PLCγ, RasGAP (proteins involved in signal transduction), ABP-1, myosin-1, spectrin (proteins found in the cytoskeleton), and neutrophil NADPH oxidase (an enzyme). In the case of a catalytic site, any catalytically active protein, such as an enzyme, can be used, particularly one whose catalytic site is known. For example, the catalytic site of the protein glutathione S-transferase (GST) can be used. Other target molecules that possess catalytic activity may include, but are not limited to, protein serine/threonine kinases, protein tyrosine kinases, serine proteases, DNA or RNA polymerases, phospholipases, GTPases, ATPases, PI-kinases, DNA methylases, metabolic enzymes, or protein glycosylases.

5.1.2. Recognition Units

By the phrase "recognition unit," is meant any molecule having a selective affinity for the functional domain of the target molecule and, preferably, having a molecular weight of up to about 20,000 daltons. In a particular embodiment of the invention, the recognition unit has a molecular weight that ranges from about 100 to about 10,000 daltons.

Accordingly, preferred recognition units of the present invention possess a molecular weight of about 100 to about 5,000 daltons, preferably from about 100 to about 2,000 daltons, and most preferably from about 500 to about 1,500 daltons. As described further below, the recognition unit of

the present invention can be a peptide, a carbohydrate, a nucleoside, an oligonucleotide, any small synthetic molecule, or a natural product. When the recognition unit is a peptide, the peptide preferably contains about 6 to about 60 amino acid residues.

When the recognition unit is a peptide, the peptide can have less than about 140 amino acid residues; preferably, the peptide has less than about 100 amino acid residues; preferably, the peptide has less than about 70 amino acid residues; preferably, the peptide has 20 to 50 amino acid residues; most preferably, the peptide has about 6 to 60 amino acid residues.

The peptide recognition units are preferably in the form of a multivalent peptide complex comprising avidin or streptavidin (optionally conjugated to a label such as alkaline phosphatase or horseradish peroxidase) and biotinylated peptides.

According to the present invention, a recognition unit (preferably in the form of a multvalent recognition unit complex) is used to screen a plurality of expression products of gene sequences containing nucleic acid sequences that are present in native RNA or DNA (*e.g.*, cDNA library, genomic library).

The step of choosing a recognition unit can be accomplished in a number of ways that are known to those of ordinary skill, including but not limited to screening cDNA libraries or random peptide libraries for a peptide that binds to the functional domain of interest. See, *e.g.*, Yu et al., 1994, Cell 76, 933-945; Sparks et al., 1994, J. Biol. Chem. 269, 23853-23856. Alternatively, a peptide or other small molecule or drug may be known to those of ordinary skill to bind to a certain target molecule and can be used. The recognition unit can even be synthesized from a lead compound, which again may be a peptide, carbohydrate, oligonucleotide, small drug molecule, or the like. The recognition unit can also be identified for use by doing searches (preferably via database) for molecules having homology for other, known

recognition unit(s) having the ability to selectively bind to the functional domain of interest.

In a specific embodiment, the step of selecting a recognition unit for use can be effected by, *e.g.*, the use of diversity libraries, such as random or combinatorial peptide or nonpeptide libraries, which can be screened for molecules that specifically bind to the functional domain of interest, *e.g.*, an SH3 domain. Many libraries are known in the art that can be used, *e.g.*, chemically synthesized libraries, recombinant (*e.g.*, phage display libraries), and *in vitro* translation-based libraries.

Examples of chemically synthesized libraries are described in Fodor et al., 1991, Science 251:767-773; Houghten et al., 1991, Nature 354:84-86; Lam et al., 1991, Nature 354:82-84; Medynski, 1994, Bio/Technology 12:709-710; Gallop et al., 1994, J. Medicinal Chemistry 37(9):1233-1251; Ohlmeyer et al., 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926; Erb et al., 1994, Proc. Natl. Acad. Sci. USA 91:11422-11426; Houghten et al., 1992, Biotechniques 13:412; Jayawickreme et al., 1994, Proc. Natl. Acad. Sci. USA 91:1614-1618; Salmon et al., 1993, Proc. Natl. Acad. Sci. USA 90:11708-11712; PCT Publication No. WO 93/20242; and Brenner and Lerner, 1992, Proc. Natl. Acad. Sci. USA 89:5381-5383.

Examples of phage display libraries are described in Scott and Smith, 1990, Science 249:386-390; Devlin et al., 1990, Science, 249:404-406; Christian, R.B., et al., 1992, J. Mol. Biol. 227:711-718); Lenstra, 1992, J. Immunol. Meth. 152:149-157; Kay et al., 1993, Gene 128:59-65; and PCT Publication No. WO 94/18318 dated August 18, 1994.

In vitro translation-based libraries include but are not limited to those described in PCT Publication No. WO 91/05058 dated April 18, 1991; and Mattheakis et al., 1994, Proc. Natl. Acad. Sci. USA 91:9022-9026.

By way of examples of nonpeptide libraries, a benzodiazepine library (*see e.g.*, Bunin et al., 1994, Proc. Natl. Acad. Sci. USA 91:4708-4712) can be adapted for use. Peptoid libraries (Simon et al., 1992, Proc. Natl. Acad. Sci.

USA 89:9367-9371) can also be used. Another example of a library that can be used, in which the amide functionalities in peptides have been permethylated to generate a chemically transformed combinatorial library, is described by Ostresh et al. (1994, Proc. Natl. Acad. Sci. USA 91:11138-11142).

The variety of non-peptide libraries that are useful in the present invention is great. For example, Ecker and Crooke, 1995, Bio/Technology 13:351-360 list benzodiazapines, hydantoins, piperazinediones, biphenyls, sugar analogs, β-mercaptoketones, arylacetic acids, acylpiperidines, benzopyrans, cubanes, xanthines, aminimides, and oxazolones as among the chemical species that form the basis of various libraries.

Non-peptide libraries can be classified broadly into two types: decorated monomers and oligomers. Decorated monomer libraies employ a relatively simple scaffold structure upon which a variety of functional groups is added. Often the scaffold will be a molecule with a known useful pharmacological activity. For example, the scaffold might be the benzodiazapine structure.

Non-peptide oligomer libraries utilize a large number of monomers that are assembled together in a ways that create new shapes that depend on the order of the monomers. Among the monomer units that have been used are carbamates, pyrrolinones, and morpholinos. Peptoids, peptide-like oligomers in which the side chain is attached to the α amino group rather than the α carbon, form the basis of another version of non-peptide oligomer libraries. The first non-peptide oligomer libraries utilized a single type of monomer and thus contained a repeating backbone. Recent libraries have utilized more than one monomer, giving the libraries added flexibility.

Screening the libraries can be accomplished by any of a variety of commonly known methods. See, *e.g.*, the following references, which disclose screening of peptide libraries: Parmley and Smith, 1989, Adv. Exp. Med. Biol. 251:215-218; Scott and Smith, 1990, Science 249:386-390;

Fowlkes et al., 1992; BioTechniques 13:422-427; Oldenburg et al., 1992, Proc. Natl. Acad. Sci. USA 89:5393-5397; Yu et al., 1994, Cell 76:933-945; Staudt et al., 1988, Science 241:577-580; Bock et al., 1992, Nature 355:564-566; Tuerk et al., 1992, Proc. Natl. Acad. Sci. USA 89:6988-6992; Ellington et al., 1992, Nature 355:850-852; U.S. Patent No. 5,096,815, U.S. Patent No. 5,223,409, and U.S. Patent No. 5,198,346, all to Ladner et al.; Rebar and Pabo, 1993, Science 263:671-673; and PCT Publication No. WO 94/18318.

In a specific embodiment, screening to identify a recognition unit can be carried out by contacting the library members with an SH3 domain immobilized on a solid phase and harvesting those library members that bind to the SH3 domain. Examples of such screening methods, termed "panning" techniques are described by way of example in Parmley and Smith, 1988, Gene 73:305-318; Fowlkes et al., 1992, BioTechniques 13:422-427; PCT Publication No. WO 94/18318; and in references cited hereinabove.

In another embodiment, the two-hybrid system for selecting interacting proteins in yeast (Fields and Song, 1989, Nature 340:245-246; Chien et al., 1991, Proc. Natl. Acad. Sci. USA 88:9578-9582) can be used to identify recognition units that specifically bind to SH3 domains.

Where the recognition unit is a peptide, the peptide can be conveniently selected from any peptide library, including random peptide libraries, combinatorial peptide libraries, or biased peptide libraries. The term "biased" is used herein to mean that the method of generating the library is manipulated so as to restrict one or more parameters that govern the diversity of the resulting collection of molecules, in this case peptides.

Thus, a truly random peptide library would generate a collection of peptides in which the probability of finding a particular amino acid at a given position of the peptide is the same for all 20 amino acids. A bias can be introduced into the library, however, by specifying, for example, that a lysine occur every fifth amino acid or that positions 4, 8,

and 9 of a decapeptide library be fixed to include only arginine. Clearly, many types of biases can be contemplated, and the present invention is not restricted to any particular bias. Furthermore, the present invention contemplates specific types of peptide libraries, such as phage-displayed peptide libraries and those that utilize a DNA construct comprising a lambda phage vector with a DNA insert.

As mentioned above, in the case of a recognition unit that is a peptide, the peptide may have about 6 to less than about 60 amino acid residues, preferably about 6 to about 25 amino acid residues, and most preferably, about 6 to about 15 amino acids. In another embodiment, a peptide recognition unit has in the range of 20-100 amino acids, or 20-50 amino acids. In the case of a bile acid receptor, for example, the recognition unit may be a bile acid, such as cholic acid or cholesterol, and may have a molecular weight of about 300 to about 600. If the functional domain relates to transcriptional control, the recognition unit may be a portion of a transcriptional factor, which may bind to a region of a gene of interest or to an RNA polymerase. The recognition unit may even be a nucleoside analog, such as cordycepin or the triphosphate thereof, capable of inhibiting RNA biosynthesis. The recognition unit may also be the carbohydrate portion of a glycoprotein, which may have a selective affinity for the asialoglycoprotein receptor, or the repeating glucan unit that exhibits a selective affinity for a cellulose binding domain or the active site of heparinase.

The selected recognition unit can be obtained by chemical synthesis or recombinant expression. It is preferably purified prior to use in screening a plurality of gene sequences.

5.1.3. Screening a Source of Polypeptides

After the recognition unit is chosen for use, the recognition unit is then contacted with a plurality of polypeptides, preferably containing a functional domain. In a particular embodiment of the invention, the plurality of

polypeptides is obtained from a polypeptide expression library. The polypeptide expression library may be obtained, in turn, from cDNA, fragmented genomic DNA, and the like. In a specific embodiment, the library that is screened is a cDNA library of total poly A+ RNA of an organism, in general, or of a particular cell or tissue type or developmental stage or disease condition or stage. The expression library may utilize a number of expression vehicles known to those of ordinary skill, including but not limited to, recombinant bacteriophage, lambda phage, M13, a recombinant plasmid or cosmid, and the like.

The plurality of polypeptides or the DNA sequences encoding same may be obtained from a variety of natural or unnatural sources, such as a procaryotic or a eucaryotic cell, either a wild type, recombinant, or mutant. In particular, the plurality of polypeptides may be endogenous to microorganisms, such as bacteria, yeast, or fungi, to a virus, to an animal (including mammals, invertebrates, reptiles, birds, and insects) or to a plant cell.

In addition, the plurality of polypeptides may be obtained from more specific sources, such as the surface coat of a virion particle, a particular cell lysate, a tissue extract, or they may be restricted to those polypeptides that are expressed on the surface of a cell membrane.

Moreover, the plurality of polypeptides may be obtained from a biological fluid, particularly from humans, including but not limited to blood, plasma, serum, urine, feces, mucus, semen, vaginal fluid, amniotic fluid, or cerebrospinal fluid. The plurality of polypeptides may even be obtained from a fermentation broth or a conditioned medium, including all the polypeptide products secreted or produced by the cells previously in the broth or medium.

The step of contacting the recognition unit with the plurality of polypeptides may be effected in a number of ways. For example, one may contemplate immobilizing the recognition unit on a solid support and bringing a solution of the plurality of polypeptides in contact with the immobilized

recognition unit. Such a procedure would be akin to an affinity chromatographic process, with the affinity matrix being comprised of the immobilized recognition unit. The polypeptides having a selective affinity for the recognition unit can then be purified by affinity selection. The nature of the solid support, process for attachment of the recognition unit to the solid support, solvent, and conditions of the affinity isolation or selection procedure would depend on the type of recognition unit in use but would be largely conventional and well known to those of ordinary skill in the art. Moreover, the valency of the recognition unit in the recognition unit complex used to screen the polypeptides is believed to affect the specificity of the screening step, and thus the valency can be chosen as appropriate in view of the desired specificity (see Sections 5.2 and 5.2.1).

Alternatively, one may also separate the plurality of polypeptides into substantially separate fractions comprising individual polypeptides. For instance, one can separate the plurality of polypeptides by gel electrophoresis, column chromatography, or like method known to those of ordinary skill for the separation of polypeptides. The individual polypeptides can also be produced by a transformed host cell in such a way as to be expressed on or about its outer surface. Individual isolates can then be "probed" by the recognition unit, optionally in the presence of an inducer should one be required for expression, to determine if any selective affinity interaction takes place between the recognition unit and the individual clone. Prior to contacting the recognition unit with each fraction comprising individual polypeptides, the polypeptides can optionally first be transferred to a solid support for additional convenience. Such a solid support may simply be a piece of filter membrane, such as one made of nitrocellulose or nylon.

In this manner, positive clones can be identified from a collection of transformed host cells of an expression library, which harbor a DNA construct encoding a polypeptide having a selective affinity for the recognition unit. The

polypeptide produced by the positive clone includes the functional domain of interest or a functional equivalent thereof. Furthermore, the amino acid sequence of the polypeptide having a selective affinity for the recognition unit can be determined directly by conventional means of amino acid sequencing, or the coding sequence of the DNA encoding the polypeptide can frequently be determined more conveniently by use of standard DNA sequencing methods. The primary sequence can then be deduced from the corresponding DNA sequence.

If the amino acid sequence is to be determined from the polypeptide itself, one may use microsequencing techniques. The sequencing technique may include mass spectroscopy.

In certain situations, it may be desirable to wash away any unbound recognition unit from a mixture of the recognition unit and the plurality of polypeptides prior to attempting to determine or to detect the presence of a selective affinity interaction (*i.e.*, the presence of a recognition unit that remains bound after the washing step) Such a wash step may be particularly desirable when the plurality of polypeptides is bound to a solid support.

As can be anticipated, the degree of selective affinities observed varies widely, generally falling in the range of about 1 nm to about 1 mM. In preferred embodiments of the present invention, the selective affinity is on the order of about 10 nM to about 100 μM, more preferably on the order of about 100 nM to about 10 μM, and most preferably on the order of about 100 nM to about 1 μM.

5.2. Specificity of Recognition Units

A particular recognition unit may have fairly *generic* selectivity for a several members (*e.g.*, three or four or more) of a "panel" of polypeptides having the domain of interest (or different versions of the domain of interest or functional equivalents of the domain of interest) or a fairly *specific* selectivity for only one or two, or possibly three,

of the polypeptides among a "panel" of same. Furthermore, multiple recognition units, each exhibiting a range of selectivities among a "panel" of polypeptides can be used to identify an increasingly comprehensive set of additional polypeptides that include the functional domain of interest.

Hence, in a population of related polypeptides, the functional domains of interest of each member may be schematically represented by a circle. See, by way of example, Figure 7A. The circle of one polypeptide may overlap with that of another polypeptide. Such overlaps may be few or numerous for each polypeptide. A particular recognition unit, A, may recognize or interact with a portion of the circle of a given polypeptide which does not overlap with any other circle. Such a recognition unit would be fairly specific to that polypeptide. On the other hand, a second recognition unit, B, may recognize a region of overlap between two or more polypeptides. Such a recognition unit would consequently be less specific than the recognition unit A and may be characterized as having a more generic specificity depending on the number of polypeptides that it recognizes or interacts with.

It should also be apparent to those of ordinary skill that any number of B-type recognition units (B_1, B_2, B_3, *etc.*) can be present, each recognizing different "panels" of polypeptides. Hence, the use of multiple recognition units provides an increasingly more exhaustive population of polypeptides, each of which exhibits a variation or evolution in the functional domain of interest present in the initial target molecule. It should also be apparent to one that the present method can be applied in an iterative fashion, such that the identification of a particular polypeptide can lead to the choice of another recognition unit. See, *e.g.*, Figure 7B. Use of this new recognition unit will lead, in turn, to the identification of other polypeptides that contain functional domains of interest that enhance the phenotypic and/or genotypic diversity of the population of "related" polypeptides.

Hence, with a given recognition unit, one may observe interaction with only one or two different polypeptides. With other recognition units, one may find three, four, or more selective interactions. In the situation in which only a single interaction is observed, it is likely, though not mandatory, that the selective affinity interaction is between the recognition unit and a replica of the initial target molecule (or a molecule very similar structurally and "functionally" to the initial target molecule).

5.2.1. Effect of the Presentation of the Recognition Unit Complex on the Specificity of the Recognition Unit-Functional Domain Interaction

The present inventors have found, unexpectedly, that the valency (*i.e.*, whether it is a monomer, dimer, tetramer, *etc.*) of the recognition unit that is used to screen an expression library or other source of polypeptides apparently has a marked effect upon which genes or polypeptides are identified from the expression library or source of polypeptides. In particular, the specificity of the recognition unit-functional domain interaction appears to be affected by the valency of the recognition unit in the screening process. By this specificity is meant the selectivity in the functional domains to which the recognition unit will bind in the screening step.

As discussed above, in one embodiment, recognition units are obtained by screening a source of recognition units, *e.g.*, a phage display library, for recognition units that bind to a particular target functional domain. Alternatively, database searches for recognition units with sequence homology to known recognition units can be employed. Of course, if a recognition unit for a particular target functional domain is already known, there is no need to screen a library or other source of recognition units; one can merely synthesize that particular recognition unit. The recognition unit, however obtained, is then used to screen an expression library or other source of polypeptides, to identify polypeptides that

the recognition unit binds to. A recognition unit that identifies only its target functional domain is a recognition unit that is completely specific. A recognition unit that identifies one or two other polypeptides that do not contain identically the target functional domain, from among a plurality of polypeptides (*e.g.*, of greater than 10^4, 10^6, or 10^8 complexity), in addition to identifying a molecule comprising its target functional domain, is very or highly specific. A recognition unit that identifies most other polypeptides present that do not contain its target functional domain, in addition to identifying its target functional domain, is a non-specific recognition unit. In between very specific recognition units and non-specific recognition units, the present inventors have discovered that there are recognition units that recognize a small number of molecules having functional domains other than their target functional domains. These recognition units are said to have generic specificity.

Thus, there is a "specificity continuum", from completely and very specific through generic to non-specific, that a recognition unit may evince. See Figure 11 for a depiction of this specificity continuum. The Applicants have discovered that a major factor influencing the specificity exhibited by a recognition unit appears to be the valency of the recognition unit in the complex used to screen the expression library.

Usually, high specificity is considered to be desirable when screening a library. High specificity is exhibited, *e.g.*, by affinity purified polyclonal antisera which, in general, are very specific. Monoclonal antibodies are also very specific. Small peptides in monovalent form, on the other hand, generally give very weak, non-specific signals when used to screen a library; thus, they are considered to be non-specific.

The present inventors have discovered that recognition units in the form of small peptides, in multivalent form, have a specificity midway between the high

specificity of antibodies and the low/non-specificity of monovalent peptides. Multivalency of the recognition unit of at least two, in a recognition unit complex used to screen the gene library, is preferred, with a multivalency of at least four more preferred, to obtain a screening wherein specificity is eased but not forfeited. In particular, a multivalent (believed to be tetravalent) recognition unit complex comprising streptavidin or avidin (preferably conjugated to a label, *e.g.*, an enzyme such as alkaline phosphatase or horseradish peroxidase, or a fluorogen, *e.g.* green fluorescent protein) and biotinylated peptide recognition units have an unexpected generic specificity. This allows such peptides to be used to screen libraries to identify classes of polypeptides containing functional domains that are similar but not identical to the peptides' target functional domains. These classes of polypeptides are identified despite the low level of homology at the amino acid level of the functional domains of the members of the classes.

In another specific embodiment, multivalent peptide recognition units may be in the form of multiple antigen peptides (MAP) (Tam, 1989, J. Imm. Meth. 124:53-61; Tam, 1988, Proc. Natl. Acad. Sci. USA 85:5409-5413). In this form, the peptide recognition unit is synthesized on a branching lysyl matrix using solid-phase peptide synthesis methods. Recognition units in the form of MAP may be prepared by methods known in the art (Tam, 1989, J. Imm. Meth. 124:53-61; Tam, 1988, Proc. Natl. Acad. Sci. USA 85:5409-5413), or, for example, by a stepwise solid-phase procedure on MAP resins (Applied Biosystems), utilizing methodology established by the manufacturer. MAP peptides may be synthesized comprising (recognition unit peptide)$_2$Lys$_1$, (recognition unit peptide)$_4$Lys$_3$, (recognition unit peptide)$_8$Lys$_6$ or more levels of branching.

The multivalent peptide recognition unit complexes may also be prepared by cross-linking the peptide to a carrier protein, *e.g.*, bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), or an enzyme, by use of known cross-linking

reagents. Such cross-linked peptide recognition units may be detected by, *e.g.*, an antibody to the carrier protein or detection of the enzymatic activity of the carrier protein.

Furthermore, the present inventors have discovered what specificity is exhibited by various types of recognition units and their complexes, *i.e.*, where these recognition units and their complexes fall in the specificity continuum. The present inventors have discovered a range of formats for presenting recognition units used to screen libraries. For example, the present inventors have determined that a peptide in the form of a bivalent fusion protein with alkaline phosphatase is very specific. The same peptide in the form of a fusion protein with the pIII protein of an M13 derived bacteriophage, expressed on the phage surface, has somewhat less, though still high, specificity. That same peptide when biotinylated in the form of a tetravalent streptavidin-alkaline phosphatase complex has generic specificity. Use of such a generically specific peptide permits the identification of a wide range of proteins from expression libraries or other sources of polypeptides, each protein containing an example of a particular functional domain.

Accordingly, the present invention provides a method of modulating the specificity of a peptide such that the peptide can be used as a recognition unit to screen a plurality of polypeptides, thus identifying polypeptides that have a functional domain. In a specific embodiment, specificity is generic so as to provide for the identification of polypeptides having a functional domain that varies in sequence from that of the target functional domain known to bind the recognition unit under conditions of high specificity. In a particular embodiment, the method comprises forming a tetravalent complex of the biotinylated peptide and streptavidin-alkaline phosphatase prior to use for screening an expression library.

5.3. Kits

The present invention is also directed to an assay kit which can be useful in the screening of drug candidates. In a particular embodiment of the present invention, an assay kit is contemplated which comprises in one or more containers (a) a polypeptide containing a functional domain of interest; and (b) a recognition unit having a selective affinity for the polypeptide. The kit optionally further comprises a detection means for determining the presence of a polypeptide-recognition unit interaction or the absence thereof.

In a specific embodiment, either the polypeptide containing the functional domain or the recognition unit is labeled. A wide range of labels can be used to advantage in the present invention, including but not limited to conjugating the recognition unit to biotin by conventional means. Alternatively, the label may comprise a fluorogen, an enzyme, an epitope, a chromogen, or a radionuclide. Preferably, the biotin is conjugated by covalent attachment to either the polypeptide or the recognition unit. The polypeptide or, preferably, the recognition unit is immobilized on a solid support. The detection means employed to detect the label will depend on the nature of the label and can be any known in the art, *e.g.*, film to detect a radionuclide; an enzyme substrate that gives rise to a detectable signal to detect the presence of an enzyme; antibody to detect the presence of an epitope, *etc.*

A further embodiment of the assay kit of the present invention includes the use of a plurality of polypeptides, each polypeptide containing a functional domain of interest. The assay kit further comprises at least one recognition unit having a selective affinity for each of the plurality of polypeptides and a detection means for determining the presence of a polypeptide-recognition unit interaction or the absence thereof.

A kit is provided that comprises, in one or more containers, a first molecule comprising an SH3 domain and a second molecule that binds to the SH3 domain, *i.e.*, a

recognition unit, where the SH3 domain is a novel SH3 domain identified by the methods of the present invention.

In a specific embodiment, the present invention provides an assay kit comprising in one or more containers:

(a) a purified polypeptide containing a functional domain of interest, in which the functional domain of is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix; and

(b) a purified recognition unit having a selective binding affinity for said functional domain in said polypeptide.

In the above assay kit, the polypeptide may comprise an amino acid sequence selected from the group consisting of SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, 221, 113-115, 118-121, 125-128, 133-139, 204-218, and 219.

In the above assay kit, the polypeptide may comprise an amino acid sequence selected from the group consisting of SEQ ID NOs:6, 14, 16, 26, 28, 34, 36, 112, 116, 117, 122-124, 129-132, and 140.

In other embodiments of the above-described assay kit, the recognition unit may be a peptide. The recognition unit may be labeled with *e.g.*, an enzyme, an epitope, a chromogen, or biotin.

In another specific embodiment, the present invention provides an assay kit comprising in containers:

(a) a plurality of purified polypeptides, each polypeptide in a separate container and each polypeptide containing a functional domain of interest in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc fingers, leucine zippers, and helix-turn-helix; and

(b) at least one recognition unit having a selective binding affinity for said functional domain in each of said plurality of polypeptides.

The present invention also provides an assay kit comprising in one or more containers:

(a) a plurality of purified polypeptides, each polypeptide in a separate container and each polypeptide containing an SH3 domain; and

(b) at least one peptide having a selective affinity for the SH3 domain in each of said plurality of polypeptides.

The present invention also provides a kit comprising a plurality of purified polypeptides comprising a functional domain of interest, each polypeptide in a separate container, and each polypeptide having a functional domain of a different sequence but capable of displaying the same binding specificity.

In the above-described kits, the polypeptides may have an amino acid sequence selected from the group consisting of: SEQ ID NOs:8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, 221.

In the above-described kits, the functional domain may be an SH3 domain.

The molecular components of the kits are preferably purified.

The kits of the present invention may be used in the methods for identifying new drug candidates and determining the specificities thereof that are described in Section 5.4.

5.4. Assays for the Identification of Potential Drug Candidates and Determining the Specificity Thereof

The present invention also provides methods for identifying potential drug candidates (and lead compounds) and determining the specificities thereof. For example, knowing that a polypeptide with a functional domain of interest and a recognition unit, *e.g.*, a binding peptide, exhibit a selective affinity for each other, one may attempt to identify a drug that can exert an effect on the polypeptide-recognition unit interaction, *e.g.*, either as an agonist or as an antagonist (inhibitor) of the interaction. With this assay, one can

screen a collection of candidate "drugs" for the one exhibiting the most desired characteristic, *e.g.*, the most efficacious in disrupting the interaction or in competing with the recognition unit for binding to the polypeptide.

Alternatively, one may utilize the different selectivities that a particular recognition unit may exhibit for different polypeptides bearing the same, similar, or functionally equivalent functional domains. Thus, one may tailor the screen to identify drug candidates that exhibit more selective activities directed to specific polypeptide-recognition unit interactions, among the "panel" of possibilities. Thus, for example, a drug candidate may be screened to identify the presence or absence of an effect on particular binding interactions, potentially leading to undesirable side effects.

Indeed, an intriguing application of the present invention is described as follows. A known antiviral agent, FIAU (a halogenated nucleoside analog), is effective at given dosages against the virus that causes hepatitis B. This compound is suspected of causing toxic side effects, however, which give rise to liver failure in certain patients to whom the drug is administered. According to the present invention, an assay is provided which can be used to develop a new generation of FIAU-derived drug that maintains its effectiveness against viral replication while reducing liver toxicity. Such an assay is provided by choosing FIAU as a recognition unit having a selective affinity for a polypeptide present in the hepatitis B virus or a cell infected with the virus. This polypeptide or family of polypeptides having the functional domain of interest is obtained by allowing the chosen recognition unit, FIAU, to come into contact with an expression library comprised of the hepatitis B virus genome and/or a cDNA expression library of infected cells, according to the methods of the present invention.

Likewise, the chosen recognition unit is allowed to come into contact with a plurality of polypeptides obtained from a sample of a human liver extract or of noninfected

hepatocytes. In this manner, a "panel" of polypeptides each of which exhibits a selective affinity for the chosen recognition unit is identified. As described above, this panel is used to determine the activities of drug (FIAU) homologs, analogs, or derivatives in terms of, say, selective inhibition of viral polypeptide-FIAU interaction versus liver polypeptide-FIAU interaction. Hence, those drug homologs, analogs, or derivatives that maintain a selective affinity for the viral polypeptide (or infected cell polypeptide) while failing to interact with or having a minimal binding affinity for liver polypeptides (and, hence, have reduced toxicity in the liver due to elimination of undesirable molecular interactions) can be identified and selected. Additional iterations of this process can be performed if so desired.

Therefore, the present invention contemplates an assay for screening a drug candidate comprising: (a) allowing at least one polypeptide comprising a functional domain of interest to come into contact with at least one recognition unit having a selective affinity for the polypeptide in the presence of an amount of a drug candidate, such that the polypeptide and the recognition unit are capable of interacting when brought into contact with one another in the absence of said drug candidate, and in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix; and (b) determining the effect, if any, of the presence of the amount of the drug candidate on the interaction of the polypeptide with the recognition unit.

In one embodiment, the effect of the drug candidate upon multiple, different interacting polypeptide-recognition unit pairs is determined in which at least some of said polypeptides have a functional domain that differs in sequence but is capable of displaying the same binding specificity as the functional domain in another of said polypeptides.

In another embodiment, at least one of said at least one polypeptide or recognition unit contains a consensus

functional domain and consensus recognition unit, respectively.

In another embodiment, the drug candidate is an inhibitor of the polypeptide-recognition unit interaction that is identified by detecting a decrease in the binding of polypeptide to recognition unit in the presence of such inhibitor.

In another embodiment, said polypeptide is a polypeptide containing an SH3 domain produced by a method comprising:

(i) screening a peptide library with an SH3 domain to obtain one or more peptides that bind the SH3 domain;

(ii) using one of the peptides from step (i) to screen a source of polypeptides to identify one or more polypeptides containing an SH3 domain;

(iii) determining the amino acid sequence of the polypeptides identified in step (ii); and

(iv) producing the one or more novel polypeptides containing an SH3 domain.

In another embodiment, said polypeptide is a polypeptide containing an SH3 domain produced by a method comprising:

(i) screening a peptide library with an SH3 domain to obtain a plurality of peptides that bind the SH3 domain;

(ii) determining a consensus sequence for the peptides obtained in step (i);

(iii) producing a peptide comprising the consensus sequence;

(iv) using the peptide comprising the consensus sequence to screen a source of polypeptides to identify one or more polypeptides containing an SH3 domain;

(v) determining the amino acid sequence of the polypeptides identified in step (iv); and

(vi) producing the one or more polypeptides containing an SH3 domain.

In a preferred embodiment, the effect of the drug candidate upon multiple, different interacting polypeptide-

recognition unit pairs is determined in which preferably at least some (*e.g.*, at least 2, 3, 4, 5, 7, or 10) of said polypeptides have functional domains that vary in sequence yet are capable of displaying the same binding specificity, *i.e.*, binding to the same recognition unit. In another specific embodiment, at least one of said polypeptides and/or recognition units contain a consensus functional domain and recognition unit, respectively (and thus are not known to be naturally expressed proteins). In one embodiment, the polypeptide is a novel polypeptide identified by the methods of the present invention. In a specific embodiment, an inhibitor of the polypeptide-recognition unit interaction is identified by detecting a decrease in the binding of polypeptide to recognition unit in the presence of such inhibitor.

A common problem in the development of new drugs is that of identifying a single, or a small number, of compounds that possess a desirable characteristic from among a background of a large number of compounds that lack that desired characteristic. This problem arises both in the testing of compounds that are natural products from plant, animal, or microbial sources and in the testing of man-made compounds. Typically, hundreds, or even thousands, of compounds are randomly screened by the use of *in vitro* assays such as those that monitor the compound's effect on some enzymatic activity, its ability to bind to a reference substance such as a receptor or other protein, or its ability to disrupt the binding between a receptor and its ligand.

The compounds which pass this original screening test are known as "lead" compounds. These lead compounds are then put through further testing, including, eventually, *in vivo* testing in animals and humans, from which the promise shown by the lead compounds in the original *in vitro* tests is either confirmed or refuted. See Remington's Pharmaceutical Sciences, 1990, A.R. Gennaro, ed., Chapter 8, pages 60-62, Mack Publishing Co., Easton, PA; Ecker and Crooke, 1995, Bio/Technology 13:351-360.

There is a continual need for new compounds to be tested in the *in vitro* assays that make up the first testing step described above. There is also a continual need for new assays by which the pharmacological activities of these compounds may be tested. It is an object of the present invention to provide such new assays to determine whether a candidate compound is capable of affecting the binding between a polypeptide containing a functional domain and a recognition unit that binds to that functional domain. In particular, it is an object of the present invention to provide polypeptides, particularly novel ones, containing functional domains and their corresponding recognition units for use in the above-described assays. The use of these polypeptides greatly expands the number of assays that may be used to screen potential drug candidates for useful pharmacological activities (as well as to identify potential drug candidates that display adverse or undesirable pharmacological activities). In one particular embodiment of the present invention, the polypeptides contain an SH3 domain.

In one embodiment of the present invention, such polypeptides are identified by a method comprising: using a recognition unit that is capable of binding to a predetermined functional domain to screen a source of polypeptides, thus identifying novel polypeptides containing the functional domain or a similar functional domain.

In a particular embodiment of the above-described method, the novel polypeptide comprises an SH3 domain and is obtained by:

(i) screening a peptide library with the SH3 domain to obtain one or more peptides that bind the SH3 domain;

(ii) using one of the peptides from step (i), preferably in the form of a multivalent complex, to screen a source of polypeptides to identify one or more novel polypeptides containing SH3 domains;

(iii) determining the amino acid sequence of the polypeptides identified in step (ii); and

(iv) producing the one or more novel polypeptides containing SH3 domains.

In another embodiment of the above-described method, the novel polypeptide containing an SH3 domain is obtained by:

(i) screening a peptide library with the SH3 domain to obtain peptides that bind the SH3 domain;

(ii) determining a consensus sequence for the peptides obtained in step (i);

(iii) producing a peptide comprising the consensus sequence;

(iv) using the peptide comprising the consensus sequence to screen a source of polypeptides to identify one or more novel polypeptides containing SH3 domains;

(v) determining the amino acid sequence of the novel polypeptides identified in step (iv); and

(vi) producing the one or more novel polypeptides containing SH3 domains.

One of ordinary skill in the art will recognize that it will not always be necessary to utilize the entire novel polypeptide containing the SH3 domain in the assays described herein. Often, a portion of the polypeptide that contains the SH3 domain will be sufficient, *e.g.*, a glutathione S-transferase (GST)-SH3 domain fusion protein. See Figure 10A and 10B for a depiction of the portions of the exemplary novel polypeptides that contain SH3 domains.

A typical assay of the present invention consists of at least the following components: (1) a molecule (*e.g.*, protein or polypeptide) comprising a functional domain; (2) a recognition unit that selectively binds to the functional domain; (3) a candidate compound, suspected of having the capacity to affect the binding between the protein containing the functional domain and the recognition unit. The assay components may further comprise (4) a means of detecting the binding of the protein comprising the functional domain and the recognition unit. Such means can be *e.g.*, a detectable label affixed to the protein comprising the functional domain, the recognition unit, or the candidate compound.

In a specific embodiment, the protein comprising the functional domain is a novel protein discovered by the methods of the present invention.

In another specific embodiment, the invention provides a method of identifying a compound that affects the binding of a molecule comprising a functional domain and a recognition unit that selectively binds to the functional domain comprising:

(a) contacting the molecule comprising the functional domain and the recognition unit under conditions conducive to binding in the presence of a candidate compound and measuring the amount of binding between the molecule and the recognition unit;

(b) comparing the amount of binding in step (a) with the amount of binding known or determined to occur between the molecule and the recognition unit in the absence of the candidate compound, where a difference in the amount of binding between step (a) and the amount of binding known or determined to occur between the molecule and the recognition unit in the absence of the candidate compound indicates that the candidate compound is a compound that affects the binding of the molecule comprising a functional domain and the recognition unit. In a specific embodiment, the molecule comprising the functional domain is a novel protein discovered by the methods of the present invention. In another specific embodiment, the functional domain is an SH3 domain.

In one embodiment, the assay comprises allowing the polypeptide containing an SH3 domain to contact a recognition unit that selectively binds to the SH3 domain in the presence and in the absence of the candidate compound under conditions such that binding of the recognition unit to the protein containing an SH3 domain will occur unless that binding is disrupted or prevented by the candidate compound. By detecting the amount of binding of the recognition unit to the protein containing an SH3 domain in the presence of the candidate compound and comparing that amount of binding to the amount of binding of the recognition unit to the protein or

polypeptide containing an SH3 domain in the absence of the candidate compound, it is possible to determine whether the candidate compound affects the binding and thus is a useful lead compound for the modulation of the activity of proteins containing the SH3 domain. The effect of the candidate compound may be to either increase or decrease the binding.

One version of an assay suitable for use in the present invention comprises binding the protein containing an SH3 domain to a solid support such as the wells of a microtiter plate. The wells contain a suitable buffer and other substances to ensure that conditions in the wells permit the binding of the protein or polypeptide containing an SH3 domain to its recognition unit. The recognition unit and a candidate compound are then added to the wells. The recognition unit is preferably labeled, *e.g.*, it might be biotinylated or labeled with a radioactive moiety, or it might be linked to an enzyme, *e.g.*, alkaline phosphatase. After a suitable period of incubation, the wells are washed to remove any unbound recognition unit and compound. If the candidate compound does not interfere with the binding of the protein or polypeptide containing an SH3 domain to the labeled recognition unit, the labeled recognition unit will bind to the protein or polypeptide containing an SH3 domain in the well. This binding can then be detected. If the candidate compound interferes with the binding of the protein or polypeptide containing an SH3 domain and the labeled recognition unit, label will not be present in the wells, or will be present to a lesser degree than is the case when compared to control wells that contain the protein or polypeptide containing an SH3 domain and the labeled recognition unit but to which no candidate compound is added. Of course, it is possible that the presence of the candidate compound will increase the binding between the protein or polypeptide containing an SH3 domain and the labeled recognition unit. Alternatively, the recognition unit can be affixed to a solid substrate during the assay. Functional

domains other than SH3 domains and their corresponding recognition units can also be used.

In a specific embodiment of the above-described method, the protein or polypeptide containing an SH3 domain is a novel protein or polypeptide containing an SH3 domain that has been identified by the methods of the present invention.

5.5. Use of Polypeptides Containing Functional Domains to Discover Polypeptides Involved in Pharmacological Activities

Using the methods of the present invention, it is possible to identify and isolate large numbers of polypeptides containing functional domains, *e.g.*, SH3 domains. Using these polypeptides, one can construct a matrix relating the polypeptides to an array of candidate drug compounds. For example, Table 1 shows such a matrix.

TABLE 1

	A	B	C	D	E	F	G	H	I	J
1										
2		X		X				X		
3										
4										
5						X				
6										
7			X					X		
8										
9	X									
10										

In Table 1, the columns headed by letters at the top of the table represent different polypeptides containing SH3 domains (preferably novel polypeptides identified by the methods of the invention). The rows numbered along the left side of the table represent recognition units with various specificity to SH3 domains. For each candidate drug compound,

a table such as Table 1 is generated from the results of binding assays. An X placed at the intersection of a particular numbered row and lettered column represents a positive assay for binding, *i.e.*, the candidate drug compound affected the binding of the recognition unit of that particular row to the SH3 domain of that particular column.

Such data as that illustrated above is used to determine whether candidate drug compounds display or are at risk of displaying desirable or undesirable physiological or pharmacological activities. For example, in Table 1, the drug compound inhibits the binding of recognition unit 2 to the SH3 domains of polypeptides B, D, and H; the compound inhibits the binding of recognition unit 5 to the SH3 domain of polypeptide F; the compound inhibits the binding of recognition unit 7 to the SH3 domains of polypeptides C and H; and the compound inhibits the binding of recognition unit 9 to the SH3 domain of polypeptide A.

If interaction with polypeptide H leads to the desirable physiological or pharmacological activity, then this drug candidate might be a good lead. However, interaction with polypeptides A, B, C, D, and F would need to be evalutated for potential side effects.

As the maps are generated and pharmacological effects observed, the maps will allow strategic assessment of the specificity necessary to obtain the desired pharmacological effect. For example, if compounds 2 and 7 are able to affect some pharmacological activity, while compounds 5 and 9 do not affect that activity, then polypeptide H is likely to be involved in that pharmacological activity. For example, if compounds 2 and 7 were both able to inhibit mast cell degranulation, while compounds 5 and 9 did not, it is likely that polypeptide H is involved in mast cell degranulation.

Accordingly, the present invention provides a method of utilizing the polypeptides comprising functional domains of the present invention in an assay to determine the participation of those polypeptides in pharmacological

activities. In a particular embodiment, the polypeptides comprise SH3 domains.

In another embodiment, the method comprises:

(a) contacting a drug candidate with a molecule comprising a functional domain under conditions conducive to binding, and detecting or measuring any specific binding that occurs; and

(b) repeating step (a) with a plurality of different molecules, each comprising a different functional domain but capable of binding to a single predetermined recognition unit under appropriate conditions.

Preferably, at least one of said molecules is a novel polypeptide identified by the methods of the present invention. In a specific embodiment, the molecules comprise the SH3 domains of Src, Abl, Cortactin, Phospholipase Cγ, Nck, Crk, p53bp2, Amphiphysin, Grb2, RasGap, or Phosphatidyl-inositol 3' kinase.

The present invention also provides a method of determining the potential pharmacological activities of a molecule comprising:

(a) contacting the molecule with a compound comprising a functional domain under conditions conducive to binding;

(b) detecting or measuring any specific binding that occurs; and

(c) repeating steps (a) and (b) with a plurality of different compounds, each compound comprising a functional domain of different sequence but capable of displaying the same binding specificity.

In a specific embodiment the functional domain is an SH3 domain.

In another embodiment, the compounds comprise the SH3 domains of Src, Abl, Cortactin, Phospholipase Cγ, Nck, Crk, p53bp2, Amphiphysin, Grb2, RasGap, or Phosphatidyl-inositol 3' kinase.

The present invention also provides a method of identifying a compound that affects the binding of a molecule

comprising a functional domain to a recognition unit that selectively binds to the functional domain comprising:

(a) contacting the molecule comprising the functional domain and the recognition unit under conditions conducive to binding in the presence of a candidate compound and measuring the amount of binding between the molecule and the recognition unit and in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix;

(b) comparing the amount of binding in step (a) with the amount of binding known or determined to occur between the molecule and the recognition unit in the absence of the candidate compound, where a difference in the amount of binding between step (a) and the amount of binding known or determined to occur between the molecule and the recognition unit in the absence of the candidate compound indicates that the candidate compound is a compound that affects the binding of the molecule comprising a functional domain and the recognition unit.

In a specific embodiment, the functional domain is an SH3 domain.

5.6. Use of More Than One Recognition Unit Simultaneously

It has been found that when screening a source of polypeptides with a recognition unit, it is possible to use more than one recognition unit at the same time. In particular, it has been found that as many as five different recognition units may be used simultaneously to screen a source of polypeptides.

In particular, when the recognition units are biotinylated peptides and the source of polypeptides is a cDNA expression library, the steps of preconjugation of the biotinylated peptides to streptavidin-alkaline phosphatase as well as the steps involved in screening the cDNA expression library may be carried out in essentially the same manner as is done when a single biotinylated peptide is used as a

recognition unit. See Section 6.1 for details. The key difference when using more than one biotinylated peptide at a time is that the peptides are combined either before or at the step where they are placed in contact with the polypeptides from which selection occurs.

In an embodiment employing a bacteriophage expression library to express the polypeptides, when the positive clones are worked up to the level of isolated plaques, the clonal bacteriophage from the isolated plaques may be tested against each of the biotinylated peptides individually, in order to determine to which of the several peptides that were used as recognition units in the primary screen the phage are actually binding.

5.7. Use of Recognition Units from Known Amino Acid Sequences

In many cases it may not be necessary to screen a collection of substances, *e.g.*, a peptide library, in order to obtain a recognition unit for a given functional domain. In the case of peptide recognition units, for example, it is sometimes possible to identify a recognition unit by inspection of known amino acid sequences. Stretches of these amino acid sequences that resemble known binding sequences for the functional domain can be synthesized and screened against a source of polypeptides in order to obtain a plurality of polypeptides comprising the given functional domain.

Prior to the disclosure of the present invention of methods of preparing recognition units having generic specificity, it would have been thought fruitless to pursue this approach. The expectation would have been that a recognition unit, chosen from published amino acid sequences as described above, would have been useful, at best, to identify a single protein containing a functional domain.

5.8. Isolation and Expression of Nucleic Acids Encoding Polypeptides Comprising a Functional Domain

In particular aspects, the invention provides amino acid sequences of polypeptides comprising functional domains, preferably human polypeptides, and fragments and derivatives thereof which comprise an antigenic determinant (*i.e.*, can be recognized by an antibody) or which are functionally active, as well as nucleic acid sequences encoding the foregoing. "Functionally active" material as used herein refers to that material displaying one or more functional activities, *e.g.*, a biological activity, antigenicity (capable of binding to an antibody) immunogenicity, or comprising a functional domain that is capable of specific binding to a recognition unit. In specific embodiments, the invention provides fragments of polypeptides comprising a functional domain consisting of at least 40 amino acids, or of at least 75 amino acids. Nucleic acids encoding the foregoing are provided. Functional fragments of at least 10 or 20 amino acids are also provided.

In other specific embodiments, the invention provides nucleotide sequences and subsequences encoding polypeptides comprising a functional domain, preferably human polypeptides, consisting of at least 25 nucleotides, at least 50 nucleotides, or at least 150 nucleotides. Nucleic acids encoding fragments of the polypeptides comprising a functional domain are provided, as well as nucleic acids complementary to and capable of hybridizing to such nucleic acids. In one embodiment, such a complementary sequence may be complementary to a cDNA sequence encoding a polypeptide comprising a functional domain of at least 25 nucleotides, or of at least 100 nucleotides. In a preferred aspect, the invention utilizes cDNA sequences encoding human polypeptides comprising a functional domain or a portion thereof.

Any eukaryotic cell can potentially serve as the nucleic acid source for the molecular cloning of polypeptides comprising a functional domain. The DNA may be obtained by standard procedures known in the art (*e.g.*, a DNA "library") by cDNA cloning, or by the cloning of genomic DNA, or

fragments thereof, purified from the desired cell (see, for example Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, 2d. Ed., Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will contain only exon sequences. Whatever the source, the gene encoding a polypeptide comprising a functional domain should be molecularly cloned into a suitable vector for propagation of the gene.

In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNAse in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once a gene encoding a particular polypeptide comprising a functional domain has been isolated from a first species, it is a routine matter to isolate the corresponding gene from another species. identification of the specific DNA fragment from another species containing the desired gene may be accomplished in a number of ways. For example, if an amount of a portion of a gene or its specific RNA from the first species, or a fragment thereof *e.g.*, the functional domain, is available and can be purified and labeled, the generated DNA fragments from another species may be screened by nucleic acid hybridization to the labeled probe (Benton, W. and Davis, R., 1977, Science 196, 180; Grunstein, M. And Hogness, D., 1975, Proc. Natl. Acad. Sci. U.S.A. 72, 3961). Those DNA fragments with substantial homology to the probe will hybridize. In a preferred embodiment, PCR using primers that hybridize to a known sequence of a gene of one species

can be used to amplify the homolog of such gene in a different species. The amplified fragment can then be isolated and inserted into an expression or cloning vector. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known restriction map if such is available. Further selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be selected which produce a protein that, *e.g.*, has similar or identical electrophoretic migration, isolectric focusing behavior, proteolytic digestion maps, *in vitro* aggregation activity ("adhesiveness") or antigenic properties as known for the particular polypeptide comprising a functional domain from the first species. If an antibody to that particular polypeptide is available, corresponding polypeptide from another species may be identified by binding of labeled antibody to the putatively polypeptide synthesizing clones, in an ELISA (enzyme-linked immunosorbent assay)-type procedure.

Genes encoding polypeptides comprising a functional domain can also be identified by mRNA selection by nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified DNA of genes encoding polypeptides comprising a functional domain of a first species. Immunoprecipitation analysis or functional assays (*e.g.*, ability to bind to a recognition unit) of the *in vitro* translation products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. In addition, specific mRNAs may be selected by adsorption of polysomes isolated from cells to immobilized antibodies specifically directed against polypeptides comprising a functional domain. A radiolabelled cDNA of a

gene encoding a polypeptide comprising a functional domain can be synthesized using the selected mRNA (from the adsorbed polysomes) as a template. The radiolabelled mRNA or cDNA may then be used as a probe to identify the DNA fragments that represent the gene encoding the polypeptide comprising a functional domain of another species from among other genomic DNA fragments. In a specific embodiment, human homologs of mouse genes are obtained by methods described above. In various embodiments, the human homolog is hybridizable to the mouse homolog under conditions of low, moderate, or high stringency. By way of example and not limitation, procedures using such conditions of low stringency are as follows (see also Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. USA 78:6789-6792): Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 μg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and reexposed to film. Other conditions of low stringency which may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations).

By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C in prehybridization mixture

containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Washing of filters is done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C for 45 min before autoradiography. Other conditions of high stringency which may be used are well known in the art.

The identified and isolated gene encoding a polypeptide comprising a functional domain can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Such vectors include, but are not limited to, bacteriophages such as lambda derivatives, or plasmids such as PBR322 or pUC plasmid derivatives. The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and gene may be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, etc., so that many copies of the gene sequence are generated.

In an alternative method, the desired gene may be identified and isolated after insertion into a suitable cloning vector in a "shot gun" approach. Enrichment for the desired gene, for example, by size fractionization, can be done before insertion into the cloning vector.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate the

isolated gene, cDNA, or synthesized DNA sequence enables generation of multiple copies of the gene. Thus, the gene may be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

The nucleic acid coding for a polypeptide comprising a functional domain of the invention can be inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. The necessary transcriptional and translational signals can also be supplied by the native gene encoding the polypeptide and/or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence. These include but are not limited to mammalian cell systems infected with virus (*e.g.*, vaccinia virus, adenovirus, *etc.*); insect cell systems infected with virus (*e.g.*, baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used.

Any of the methods previously described for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of nucleic acid sequence encoding a protein or peptide fragment may be regulated by a second nucleic acid sequence so that the protein or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of a protein may be controlled by any promoter/enhancer element known in the art. Promoters which

may be used to control gene expression include, but are not limited to, the SV40 early promoter region (Benoist and Chambon, 1981, Nature 290, 304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, Cell 22, 787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78, 1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296, 39-42); prokaryotic expression vectors such as the β-lactamase promoter (Villa-Kamaroff, et al., 1978, Proc. Natl. Acad. Sci. U.S.A. 75, 3727-3731), or the tac promoter (DeBoer, et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80, 21-25); see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242, 74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., Nature 303, 209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, et al., 1981, Nucl. Acids Res. 9, 2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310, 115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38, 639-646; Ornitz et al., 1986, Cold Spring Harbor Symp. Quant. Biol. 50, 399-409; MacDonald, 1987, Hepatology 7, 425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315, 115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, Cell 38, 647-658; Adames et al., 1985, Nature 318, 533-538; Alexander et al., 1987, Mol. Cell. Biol. 7, 1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45, 485-495), albumin gene control region which is

active in liver (Pinkert et al., 1987, Genes and Devel. 1, 268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5, 1639-1648; Hammer et al., 1987, Science 235, 53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1, 161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315, 338-340; Kollias et al., 1986, Cell 46, 89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48, 703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314, 283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234, 1372-1378).

Expression vectors containing inserts of genes encoding polypeptides comprising a functional domain can be identified by three general approaches: (a) nucleic acid hybridization, (b) presence or absence of "marker" gene functions, and (c) expression of inserted sequences. In the first approach, the presence of a foreign gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to the inserted gene. In the second approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, transformation phenotype, occlusion body formation in baculovirus, etc.) caused by the insertion of foreign genes in the vector. For example, if the gene encoding a polypeptide comprising a functional domain is inserted within the marker gene sequence of the vector, recombinants containing the gene can be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying the foreign gene product expressed by the recombinant. Such assays can be based, for example, on the physical or

functional properties of the gene product in in vitro assay systems, e.g., ability to bind to recognition units.

Once a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As previously explained, the expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (e.g., lambda), and plasmid and cosmid DNA vectors, to name but a few.

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the protein may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (e.g., glycosylation, cleavage) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce an unglycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can be used to ensure "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may effect processing reactions such as proteolytic cleavages to different extents.

In other specific embodiments, polypeptides comprising a functional domain, or fragments, analogs, or derivatives thereof may be expressed as a fusion, or chimeric protein product (comprising the polypeptide, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence (of a different protein)). Such a chimeric

product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper reading frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer.

5.8.1 Identification and Purification of the Expressed Gene Product

Once a recombinant which expresses the gene sequence encoding a polypeptide comprising a functional domain is identified, the gene product may be analyzed. This can be achieved by assays based on the physical or functional properties of the product, including radioactive labelling of the product followed by analysis by gel electrophoresis.

Once the polypeptide comprising a functional domain is identified, it may be isolated and purified by standard methods including chromatography (*e.g.*, ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. The functional properties may be evaluated using any suitable assay, including, but not limited to, binding to a recognition unit.

5.9 Derivatives and Analogs of Polypeptides Comprising a Functional Domain

The invention further provides derivatives (including but not limited to fragments) and analogs of polypeptides that are functionally active, *e.g.*, comprising a functional domain. In a specific embodiment, the derivative or analog is functionally active, *i.e.*, capable of exhibiting one or more functional activities associated with a full-length, wild-type polypeptide, *e.g.*, binding to a recognition unit. As one example, such derivatives or analogs may have the antigenicity of the full-length polypeptide.

In particular, derivatives can be made by altering gene sequences encoding polypeptides comprising a functional domain by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences which encode substantially the same amino acid sequence as a gene encoding a polypeptide comprising a functional domain may be used in the practice of the present invention. These include but are not limited to nucleotide sequences comprising all or portions of such genes which are altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of a polypeptide comprising a functional domain including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

Derivatives or analogs of genes encoding polypeptides comprising a functional domain include but are not limited to those polypeptides which are substantially homologous to the genes or fragments thereof, or whose

encoding nucleic acid is capable of hybridizing to a nucleic acid sequence of the genes.

The derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned gene sequence can be modified by any of numerous strategies known in the art (Maniatis, T., 1989, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. PCR primers can be constructed so as to introduce desired sequence changes during PCR amplification of a nucleic acid encoding the desired polypeptide. In the production of the gene encoding a derivative or analog, care should be taken to ensure that the modified gene remains within the same translational reading frame, uninterrupted by translational stop signals, in the gene region where the desired activity is encoded.

Additionally, the sequence of the genes encoding polypeptides comprising a functional domain can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, including but not limited to, *in vitro* site-directed mutagenesis (Hutchinson, C., et al., 1978, J. Biol. Chem 253:6551), use of TAB® linkers (Pharmacia), etc.

Manipulations of the sequence may also be made at the protein level. Included within the scope of the invention are protein fragments or other derivatives or analogs which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other

cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; etc.

In addition, analogs and derivatives can be chemically synthesized. For example, a peptide corresponding to a portion of a polypeptide comprising a functional domain can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids, α-amino isobutyric acid, 4-aminobutyric acid, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, designer amino acids such as β-methyl amino acids, Cα-methyl amino acids, and Nα-methyl amino acids.

5.10 Antibodies to Polypeptides Comprising a Functional Domain

According to one embodiment, the invention provides antibodies and fragments thereof containing the binding domain thereof, directed against polypeptides comprising a functional domain. Accordingly, polypeptides comprising a functional domain, fragments or analogs or derivatives thereof, in particular, may be used as immunogens to generate antibodies against such polypeptides, fragments or analogs or derivatives. Such antibodies can be polyclonal, monoclonal, chimeric, single chain, Fab fragments, or from an Fab expression library. In a specific embodiment, antibodies specific to the functional domain of a polypeptide comprising a functional domain may be prepared.

Various procedures known in the art may be used for the production of polyclonal antibodies. In a particular

embodiment, rabbit polyclonal antibodies to an epitope of a polypeptide comprising a functional domain, or a subsequence thereof, can be obtained. For the production of antibody, various host animals can be immunized by injection with the native polypeptide comprising a functional domain, or a synthetic version, or fragment thereof, including but not limited to rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhold limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum.

For preparation of monoclonal antibodies, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256, 495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4, 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Antibody fragments which contain the idiotype (binding domain) of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragment, and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, e.g. ELISA (enzyme-linked immunosorbent assay).

6. EXAMPLES

6.1. Identification of Genes from cDNA Expression Libraries

A study was initiated to determine whether peptide recognition units could recognize functional domains that are the same as or similar to their target functional domain but that are contained in proteins other than the protein containing their target functional domain. Such "functional" screens, using recognition units of relatively small size, were not previously known and were difficult to develop because of the low degree of sequence homology among functional domain-containing proteins. Thus, for example, an oligonucleotide probe could not be designed with any degree of confidence based on the low degree of homology of primary sequences of SH3 domains.

Using SH3 domain-binding peptides from combinatorial peptide libraries as recognition units, we screened a series of mouse and human cDNA expression libraries. We found that 69 of the 74 clones isolated from the libraries encoded at least one SH3 domain. These clones represent more than 18 different SH3 domain-containing proteins, of which more than 10 have not been described previously.

The initial recognition unit chosen was a Src SH3 domain-binding peptide (termed pSrcCII) isolated from a phage-displayed random peptide library (Sparks et al., 1994, J. Biol. Chem. 269:23853-23856). pSrcCII was (biotin-SGSGGILAPPVPPRNTR-NH_2) (SEQ ID NO:1). pSrcCII was synthesized by standard FMOC chemistry, purified by HPLC, and its structure was confirmed by mass spectrometry and amino acid analysis. To form multivalent complexes, 50 pmol biotinylated pSrcCII peptide was incubated with 2 μg streptavidin-alkaline phosphatase (SA-AP) (for a biotin:biotin-binding site ratio of 1:1). Excess biotin-binding sites were blocked by addition of 500 pmol biotin. Alternatively, 31.2 μl of 1 mg/ml SA-AP could have been incubated with 15 μl of 0.1 mM biotinylated peptide for 30 min at 4 °C. Ten μl of 0.1 mM biotin would

then be added, and the solution incubated for an additional 15 min.

A λEX*lox* mouse 16 day embryo cDNA expression library was obtained from Novagen (Madison, WI). The cDNA library was screened according to published protocols (Young and Davis, 1983, Proc. Natl. Acad. Sci. USA 80:1194-1198). The library was plated at an initial density of 30,000 plaques/100 mm petri plate as follows. A library aliquot was diluted 1:1000 in SM (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris HCl pH 7.5, 0.01% gelatin). Three μl of diluted phage were added to 1.5 ml each of SM, 10 mM $CaCl_2/MgCl_2$, and an overnight culture of BL21(DE3)pLysE *E. coli* cells. BL21 overnight cultures were grown in 2xYT medium (1.6% tryptone, 1% yeast extract, and 0.5% NaCl) supplemented with 10 mM $MgSO_4$, 0.2 % maltose, and 25μg/ml chloramphenicol. This mixture was incubated 20 min at 37°C, after which 300 μl were plated on each of 14 2xYT agar plates in 3 ml 0.8% 2xYT top agarose containing 25 μg/ml chloramphenicol. Plaques were allowed to form for 6 hours at 37°C, after which isopropyl-β-D-thiogalactopyranoside (IPTG)-soaked filters were applied. After an additional eight hours' incubation at 37°C, the filters were marked, removed from the plates, and washed three times with phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4), 0.1% Triton X-100. The filters were blocked for 1 hour in PBS, 2% bovine serum albumin (blocking solution) and subsequently incubated overnight at 4°C with fresh blocking solution plus streptavidin-alkaline phosphatase (SA-AP) complexed peptide. Approximately 1 μg SA-AP complexed with peptide in 1 ml blocking solution was used for each filter. The filters were then subjected to four 15 minute washes with PBS, 0.1% Triton X-100. Bound SA-AP-peptide complexes were detected by incubation with 44 ml nitroblue tetrazolium chloride (NBT, 75 mg/ml in 70% dimethylformamide) and 33 ml of 5-bromo-4-chloro-3-indoyl-phosphate-p-toluidine salt (BCIP 50 mg/ml in dimethylformamide) in 10 ml of alkaline phosphatase buffer (0.1 M Tris-HCl, pH 9.4, 0.1 M NaCl, 50 mM $MgCl_2$); the signals were robust, often evident within a few minutes.

Positive plaques were cored with a Pasteur pipet and placed in 1 ml SM with a drop of chloroform. Lambda phage particles are structurally resistant to chloroform, which serves as a bacteriocidal agent. These cores were allowed to diffuse into solution for at least 1 hr before subsequent platings. Phage from cores were plated in 100 µl each of SM, 10 mM $CaCl_2/MgCl_2$, and an overnight culture of BL21 (DE3) pLySE cells. Phage were plated with the intention of reducing the number of plaque forming units (pfu)/plate by roughly a factor of 10 with each screen (*i.e.*, 3 x 10^4 in the primary screen, 3 X 10^3 in the secondary, and so on). This was accomplished by diluting cores 1:1000 and plating 1-10 µl/plate. Four screens were generally required to obtain isolated plaques.

Plasmids were rescued from the λEX*lox* phage by cre-mediated excision in BM25.8 *E. coli* cells. For each clone, 5 µl of a 1:100 dilution of phage were added to a solution containing 100 µl SM and 100 µl of an overnight culture of BM25.8 cells (grown in 2xYT media supplemented with 10 mM $MgSO_4$, 0.2 % maltose, 34 µg/ml chloramphenicol, and 50 µg/ml kanamycin). After 30 minutes at 37 °C, 100 µl of this solution were spread on an LB amp agarose plate and incubated overnight at 37 °C. A single colony from each plate was used to inoculate 3 ml of 2xYT/amp and incubated overnight. Plasmid DNA was purified from the overnight culture using Promega Wizard Miniprep DNA purification kits (Promega, Madison, WI), extracted with an equal volume of phenol/chloroform followed by chloroform alone, and ethanol precipitated. This plasmid DNA was used to transform chemical-competent DH5α cells. Three colonies from each transformation were used to inoculate 3 ml cultures; DNA was purified as described above. Approximately, 1/20 of each individually purified DNA sample from transformed cells was digested with EcoR1 and HindIII and examined by electrophoresis on a 1% agarose gel to determine insert size and DNA quality. One DNA prep for each clone was either sequenced manually using the dideoxy method or by an automated technique that uses fluorescent dideoxynucleotide terminators.

The T7 gene 10 primer located approximately 40 bp upstream of the EcoR1 restriction site was used conveniently in both cases.

Approximately 100 of $1X10^6$ plaques in the primary screen of the λEX*lox* 16 day mouse embryo cDNA expression library exhibited significant pSrcCII-binding activity. Figure 5 is representative of filters from primary and tertiary screens. Of the eighteen positive clones that were isolated and sequenced, all were found to encode proteins with SH3 domains, although several clones appeared to be siblings or to originate from the same mRNA. Thus, the pSrcCII screen resulted in the identification of cDNAs encoding nine distinct SH3 domain-containing proteins (see Figure 9). The sequences of these proteins were compared to the sequences in GenBank with the computer program BLAST. Three of these proteins corresponded to entries in GenBank. SH3P1 appears to be the murine homologue of p53bp2, a p53-binding protein, p53bp2 (Iwabuchi et al., 1994, Proc. Natl. Acad. Sci. USA 91:6098-6102); SH3P6 resembles human MLN50, a gene amplified in some breast carcinomas (Tomasetto et al., 1995, Genomics 28:367-376); and SH3P5 is Cortactin, a protein implicated in cytoskeletal organization (Wu and Parsons, 1993, J. Cell Biol. 120:1417-1426). Six of the clones did not match entries in GenBank, indicating that the present invention can be used to identify novel SH3 domain-containing proteins. Of these novel proteins, SH3P2 contains three ankyrin repeats and a proline-rich region flanking its SH3 domain; SH3P7 and SH3P9 contain sequences related to regions in the proteins drebrin (Ishikawa et al., 1994, J. Biol. Chem. 269:29928-29933) and amphiphysin (David et al., 1994, FEBS Lett. 351:73-79), respectively. Finally, the novel proteins SH3P4 and SH3P8, although not similar to any known proteins, are highly related (89% amino acid similarity) to one another.

The present invention can be used as part of an iterative process in which a recognition unit is used to identify proteins containing functional domains which are, in turn, used to derive additional recognition units for

subsequent screens. For example, to define the binding specificity of these newly cloned SH3 domains, they can be overexpressed as glutathione S-transferase (GST)-fusion proteins in bacteria, which, in turn, can be used to screen a random peptide library in order to obtain recognition units which, in turn, can be used to screen cDNA libraries in order to obtain still more novel proteins containing SH3 domains.

The recognition unit binding preferences of two of the SH3 domains isolated in the pSrcCII screen described above (p53bp2 and Cortactin) have been described (Sparks et al., 1996, Proc. Natl. Acad. Sci. USA 93:1540-1544. Each of these SH3 domains recognizes recognition unit motifs related to, yet distinct from, the pSrcCII sequence. We used a synthetic peptide (pCort) containing the Cortactin SH3 recognition unit motif to screen the mouse embryo cDNA expression library. pCort was (biotin-SGSGSRLTPQSKPPLPPKPSWVSR-NH_2) (SEQ ID NO:2). pCort was prepared and complexed with SA-AP as above for pSrcCII. Screening of the mouse embryo library with pCort was done as above for pSrcCII.

Twenty six clones, of varying signal strength, were isolated and twenty-one were found to encode SH3 domain containing proteins. The pCort screen yielded genes corresponding to nine distinct SH3 domain-containing proteins (see Figure 9), four of which corresponded to entries in GenBank. SH3P5 and SH3P6 are Cortactin and MLN50, discussed above; SH3P10 matched SPY75/HS1, a protein involved in IgE signaling (Fukamachi et al., 1994, J. Immunol. 152:642-652); and SH3P11 is Crk, an SH2 domain and SH3 domain-containing adaptor molecule (Knudsen et al., 1994, J. Biol. Chem. 269:32781-32787). The five novel transcripts encode SH3P7, SH3P8, and SH3P9, discussed above; SH3P13, an additional member of the SH3P4/SH3P8 family; and SH3P12, a protein with three SH3 domains and a region sharing significant sequence similarity with the peptide hormone sorbin (Vagen-Descroiz M. et al., 1991, Eur. J. Biochem. 201:53-50).

Interestingly, the output from the pCort screen only partially overlapped with that of the pSrcCII screen: four of

the nine SH3-containing proteins isolated with pCort were not identified with pSrcCII. In addition, SH3P9, the protein identified most frequently (50%) in the pSrcCII screen was isolated at a much lower frequency (7%) with the pCort probe. Thus, different recognition units can be used to identify distinct sets of SH3 domains.

In addition to possessing at least one SH3 domain, a prominent characteristic of the proteins identified in the pSrcCII and pCort screens is the position of the SH3 domain within the proteins: twelve of thirteen proteins possess SH3 domains near their C-termini. Although pSrcCII binds well to the Src SH3 domain (Figure 8), Src (whose SH3 domain occurs near the N-terminus) was not identified in the pSrcCII screen. We suspect the bias was a consequence of the fact that the mouse embryo cDNA library was constructed using oligo-dT-primed cDNA. Alternatively, it may be that the mRNA used to prepare the library contained very little, or no, Src transcripts.

A variant of the pSrcCII peptide (T12SRC.1) was used to probe a λgt22a human prostate cancer cell line cDNA library primed with oligo-dT and a λgt11 human bone marrow library primed with random and oligo-dT primers. T12SRC.1 was (biotin-GILAPPVPPRNTR-NH_2) (SEQ ID NO:3). T12SRC.1 was used in the initial screens together with the peptide T12SRC.4. T12SRC.4 was (biotin-VLKRPLPIPPVTR-NH_2) (SEQ ID NO:4). The λgt22a human prostate cancer cell line cDNA library was made from the LNCaP prostate cancer cell line by using standard methods, *i.e.*, the Superscript Lambda system for cDNA synthesis and cloning (Bethesda Research Laboratories, Gaithersburg, MD). The λgt11 human bone marrow cDNA expression library was obtained from Clonetch (Palo Alto, CA). The human libraries were screened and positive clones isolated as described above for the mouse 16 day embryo cDNA library, except that cDNA inserts of the λgt11 and λgt22a phage were amplified by PCR rather than being rescued by cre-mediated excision. Of the $1.2X10^7$ λcDNA clones screened from these libraries, 30 exhibited detectable pSrcCII-binding activity.

Analysis of the positive clones revealed that they each encoded at least one SH3 domain, and that they originated from a total of six different transcripts (Figure 9). Three of these encode proteins possessing non-C-terminal SH3 domains, indicating that the present invention can be used to identify active domains regardless of their position within a protein. Of the six proteins identified, only three matched GenBank entries. SH3P15 and SH3P16 are Fyn (Kawakami et al., 1988, Proc. Natl. Acad. Sci. USA 85:3870-3874 and Lyn (Yamanashi et al., 1987, Mol. Cell. Biol. 7:237-243), respectively, two Src-family members possessing SH3 domains with ligand preferences similar to that of the Src SH3 domain (Rickles, 1994, EMBO J. 13:5598-5604); and SH3P14 appears to be the human homologue of murine H74, a protein of unknown function. The three remaining proteins did not match entries in GenBank and include the human homolog of SH3P9, described above, and SH3P17 and SH3P18, fragments of two related (85% amino acid similarity) adaptor-like proteins comprised of at least four and three SH3 domains, respectively.

Examination of the primary sequences of the SH3 domains identified in this work reveals several interesting features (see Figure 10). Positions important for ligand binding by the Src SH3 domain (Feng et al., 1994, Science 266:1241-1247; Lescure et al., 1992, J. Mol. Biol. 228:387-94) and essential for SH3 function in Grb2/Sem5 are conserved (Clark et al., 1992, Nature 356:340-344). In addition, the two gaps in the sequence alignment shown in Figure 10 correspond to regions of length variation observed among previously characterized SH3 domains. Surprisingly, the SH3 domains identified in this work are not significantly more similar to one another than they are to other known SH3 domains, with the exception of the mouse and human forms of SH3P9 and SH3P14 which are 100% and 83% identical, respectively. This result indicates that SH3 domains can vary widely in primary structure and still bind proline-rich peptide recognition units selectively.

6.1.1. Nucleotide and Corresponding Amino Acid Sequences of Genes Identified from cDNA Expression Libraries

The nucleotide sequences of SH3P1, SH3P2, SH3P3, SH3P4, SH3P5, SH3P6, SH3P7, SH3P8, SH3P9, SH3P10, SH3P11, SH3P12, SH3P13, and SH3P14, the mouse genes identified by screening the 16 day mouse embryo cDNA expression library with the peptides pSrcII and pCort, are shown in Figures 18, 20, 22, 24, 26, 28, 30, 32, 34, 38, 40, 42A and B, 44, and 46A and B, respectively. The corresponding amino acid sequences of the mouse genes SH3P1, SH3P2, SH3P3, SH3P4, SH3P5, SH3P6, SH3P7, SH3P8, SH3P9, SH3P10, SH3P11, SH3P12, SH3P13, and SH3P14 are shown in Figures 19, 21, 23, 25, 27, 29, 31, 33, 35, 39, 41, 43, 45, and 47, respectively.

The nucleotide sequences of SH3P9, SH3P14, SH3P17, and SH3P18, human genes identified by screening the human bone marrow and human prostate cancer cDNA expression libraries with the peptide T12SRC.1, are shown in Figures 36, 48, 50, and 52, respectively. The corresponding amino acid sequences of the human genes SH3P9, SH3P14, SH3P17, and SH3P18 are shown in Figures 37, 49, 51, and 53, respectively.

Two genes, SH3P9 and SH3P14, were isolated from both mouse and human libraries.

The sequences of SH3P15 and SH3P16 are not shown. SH3P15 is Lyn and SH3P16 is Fyn.

Figure 54 shows the nucleotide sequence of clone 55, a novel human gene identified and isolated from a human bone marrow cDNA library (described in Section 6.1) using as recognition units a mixture of T12SRC.4 and pCort (described in Section 6.1) and the methods described in Section 6.1.

Figure 55 shows the amino acid sequence of clone 55.

Figure 56 shows the nucleotide sequence of clone 56, a novel human gene identified and isolated from a human bone marrow cDNA library (described in Section 6.1) using as recognition units a mixture of T12SRC.4 and pCort (described in Section 6.1) and the methods described in Section 6.1.

Figure 57 shows the amino acid sequence of clone 56.

Figure 58A shows the nucleotide sequence from position 1-1720 and Figure 58B shows the nucleotide sequence from position 1720-2873 of clone 65, a novel human gene identified and isolated from a human bone marrow cDNA library (described in Section 6.1) using as recognition units a mixture of P53BP2.Con and Nck1.Con3 and the methods described in Section 6.1. P53BP2.Con and Nck1.Con3 are peptides, the amino acid sequences of which are biotin-SFAAPARPPVPPRKSRPGG-NH_2 (SEQ ID NO:201) and biotin-SFSFPPLPPAPGG-NH_2 (SEQ ID NO:202), respectively. The sequences of P53BP2.Con and Nck1.Con3 are consensus sequences of recognition units that bind to the SH3 domains of p53bp2 and Nck, respectively.

Figure 59 shows the amino acid sequence of clone 65.

Figure 60 shows the nucleotide sequence of clone 34, a novel human gene identified and isolated from a human prostate cancer cDNA library (described in Section 6.1) using as recognition units a mixture of T12SRC.1 and T12SRC.4 (described in Section 6.1) and the methods described in Section 6.1.

Figures 61A and 61B show the amino acid sequence of clone 34.

Figure 62 shows the nucleotide sequence of clone 41, a novel human gene identified and isolated from a human bone marrow cDNA library (described in Section 6.1) using as recognition units a mixture of PXXP.NCK.S1/4 and PXXP.ABL.G1/2M and the methods described in Section 6.1. PXXP.NCK.S1/4 and PXXP.ABL.G1/2M are peptides, the amino acid sequences of which are biotin-SRSLSEVSPKPPIRSVSLSR-NH_2 (SEQ ID NO:222) and biotin-SRPPRWSPPPVPLPTSLDSR-NH_2 (SEQ ID NO:223), respectively. PXXP.NCK.S1/4 and PXXP.ABL.G1/2M bind to the SH3 domains of Nck and Abl, respectively

Figures 63A and 63B show the amino acid sequence of clone 41.

Figure 64 shows the nucleotide sequence of clone 53, a novel human gene identified and isolated from a human prostate cancer cDNA library (described in Section 6.1) using

as recognition units a mixture of PXXP.NCK.S1/4 and PXXP.ABL.G1/2M and the methods described in Section 6.1.

Figures 65A and 65B show the amino acid sequence of clone 53.

Figures 66A and 66B show the nucleotide and amino acid sequence of clone 5, a novel human gene identified and isolated from a HELA cell cDNA library using as recognition units a mixture of T12SRC.1 and T12SRC.4 (described in Section 6.1) and the methods described in Section 6.1.

6.2. Use of Peptides Resembling SH3 Domain Binding Sequences as Recognition Units

We inspected a number of published amino acid sequences and identified proline-rich stretches of amino acids that resembled consensus SH3 domain binding sequences. Peptides comprising these proline-rich sequences were synthesized and tested by the methods of the present invention for their ability to specifically bind to the novel SH3 domains described in Sections 6.1 and 6.1.1. Purified SH3 domain-containing clones were spotted on a lawn of Y1090 host cells, grown for an appropriate amount of time, and plaque filter lifts were screened with biotinylated peptides complexed with streptavidin-alkaline phosphatase as described in Section 6.1.

The results are shown in Figures 12 and 13. As can be seen, in many cases the synthesized peptides were able to bind to the novel SH3 domains. This indicates that those synthesized peptides could have been used to identify those novel SH3 domains from sources of polypeptides.

6.3. Valency of Peptide Recognition Units Affects Specificity of Recognition Units

6.3.1 Preconjugation of Peptide Recognition Units with Streptavidin-Alkaline Phosphatase Increases Affinity of the Recognition Units for Targets

As a preliminary test of the effect of the valency of peptide recognition units on the ability of those

recognition units to be used as probes to detect SH3 domains, biotinylated peptides that had been previously shown to bind the SH3 domains of either Src or Abl were tested for their ability to bind their respective SH3 domain when either preconjugated with streptavidin-alkaline phosphatase (SA-AP) or not so preconjugated. GST-SrcSH3 and GST-AblSH3 fusion proteins (produced as described in Sparks et al., 1994, J. Biol. Chem. 269:23853-23856) were resolved by 10% SDS-PAGE and transferred to an Immobilon D nylon membranes (Millipore, New Bedford, MA). The membranes were incubated in blocking solution for 1 hr at 25 °C and then incubated overnight at 4 °C with either biotinylated Src SH3 domain or biotinylated Abl SH3 domain binding peptides in either multivalent (SA-AP) or monovalent format. The filters were washed three times (15 min each wash) in PBS/T and incubated with NBT and BCIP for color development. See Section 6.1 for further details of the detection process.

The results are shown in Figure 14. In panels A, the biotinylated peptides were preconjugated with SA-AP and then allowed to bind to the immobilized SH3 domains. Preconjugation was as described in Section 6.1. In panels B, the peptides were first allowed to bind to the immobilized SH3 domains and then the bound peptides were detected by adding SA-AP. In both cases, color development was as in Section 6.1. The sequences of the peptides used were: Biotin-SGSGGILAPPVPPRNTR (SEQ ID NO:1) for the Src specific peptide and Biotin-SGSGSRPPRWSPPPVPLPTSLDSR (SEQ ID NO:41) for the Abl specific peptide. The results shown in Figure 14 demonstrate that preconjugation with SA-AP dramatically increases the strength of the signal detected.

6.3.2. Preconjugation of Peptide Recognition Units with Streptavidin-Alkaline Phosphatase Results in Recognition of a Variety of SH3 Domains

Two μg of each of a panel of GST-SH3 domain fusion proteins were transferred to Immobilon D nylon membranes (Millipore, New Bedford, MA) using a dot-blot apparatus.

Biotinylated Src, Abl, or Cortactin SH3 domain-binding peptides were preconjugated to SA-AP and incubated with the filter; an alkline-phophatase driven color reaction was used to detect peptide binding. The panel of immobilized proteins was also reacted with a polyclonal anti-GST antibody (Pharmacia, Piscataway, NJ). Sequences of the Src, Abl, and Cortactin-binding peptides were Biotin-SGSGVLKRPLPIPPVTR (SEQ ID NO:42), Biotin-SGSGSRPPRWSPPPVPLPTSLDSR (SEQ ID NO:41), and Biotin-SGSGSRLGEFSKPPIPQKPTWMSR (SEQ ID NO:43), respectively.

As can be seen from the results shown in Figure 15, the preconjugated biotinylated peptides recognized not only their original target SH3 domains, but related domains as well. The Src peptide recognized the SH3 domains of Yes and Cortactin as well as the SH3 domain of Src; the Abl peptide recognized the Cortactin SH3 domain as well as the Abl SH3 domain; and the Cortactin peptide recognized Src, Yes, Abl, Crk, and the C terminal Grb2 SH3 domains as well as recognizing the Cortactin SH3 domain.

The above experiment was performed utilizing SH3 domains that had been immobilized on nylon membranes. The following demonstrates that preconjugation with streptavidin also permits peptide recognition units to recognize a variety of SH3 domains when those domains are immobilized in the wells of a microtiter plate.

Five different peptide recognition units (pAbl, pPLC, pCrk, pSrcCI, pSrcCII) were tested in either multivalent or monovalent format for their ability to bind to seven different SH3 domains (Src, Abl, PLCγ, Crk, Cortactin, Grb2N, Grb2C) in an ELISA. The sequences of these peptides were as follows: pAbl, SGSGSRPPRWSPPPVPLPTSLDSR (SEQ ID NO:41); pPLC, SGSGSMPPPVPPRPPGTLGG (SEQ ID NO:66); pCrk, SGSGNYVNALPPGPPLPAKNGG (SEQ ID NO:67); pSrcCI, SGSGVLKRPLPIPPVTR (SEQ ID NO:42); pSrcCII, SGSGGILAPPVPPRNTR (SEQ ID NO:1). These peptides were biotinylated as in Section 6.1.

The SH3 domains were produced as GST-SH3 fusion proteins as described in Sparks et al., 1994, J. Biol. Chem.

269:23853-23856. Their purity and concentration were confirmed by SDS-PAGE and Bradford protein assays, respectively. The GST-SH3 fusion proteins were immobilized in the wells of microtiter plates as follows: Two micrograms of each GST-SH3 fusion protein were incubated in wells of a flat bottom enzyme linked immunoabsorbent assay (ELISA) microtiter plate (Costar, Cambridge, MA) in 100 mM $NaHCO_3$ for 1hr 25 °C. One volume of SuperBlock blocking buffer (Pierce Chemical Co., Rockford, IL) was added to each well and incubated for an additional 30 min. Plates were washed three times with PBS/0.1% Tween-20/0.1% bovine serum albumin (BSA). Immobilized proteins were detected with SH3 domain-binding peptides in multivalent or monovalent formats using streptavidin-horseradish peroxidase (SA-HRP; Sigma Chemical Co., St. Louis, MO). For complexation of the biotinylated peptides and SA-HRP, peptide and SA-HRP concentrations were as described for SA-AP complexation in Section 6.1, but all incubations and washes were in PBS/0.1% Tween-20/0.1% BSA. Plates were washed five times before colorimetric reaction and before the addition of SA-HRP (monovalent format). The amount of bound SA-HRP was evaluated with the addition of 100 µl horseradish peroxidase substrate [2',2'-Azino-Bis 3-Ethylbenzthiazoline-6-Sulfonic Acid (ABTS), 0.05 % hydrogen peroxide, 50 mM sodium citrate, pH 5.0]. After 5-30 minutes of reaction time, the optical densities (OD) of the microtiter plate wells were measured with a microtiter plate scanner (Molecular Devices, Sunnyvale, CA) set for 405 nm wavelength. The results are shown in Figure 8. From Figure 8 it can be seen that the tetravalent (complexed) peptides display both increased affinity and broadened specificity toward SH3 targets. Binding of complexed peptides was, however, still restricted to SH3 domains; the complexes bind to neither GST (Figure 8) nor other unrelated proteins (data not shown). Thus, precomplexation with SA-AP decreases the specificity of the peptide recognition units but does not make the peptides non-specific. Rather, the peptides, when precomplexed,

recognize a variety of SH3 domains in addition to their target domains.

6.3.3. Preconjugation of Peptide Recognition Units with Streptavidin-Alkaline Phosphatase Results in Recognition of a Variety of Expressed cDNA Clones

Lambda phage clones of genes containing a variety of SH3 domains were isolated from screens of a 16 day mouse embryo cDNA expression library (Novagen, Madison, WI). For a description of the isolation of these cDNA clones, see Section 6.1. Phage particles corresponding to individual lambda phage cDNA recombinants were spotted onto 2xYT-1.5 % agar petri plates onto which had been poured 3 ml of 2xYT-0.8 % agarose with 100 μl of a BL21(DE3)pLysE *E. coli* culture grown overnight. After a 6 hr incubation at 37 °C, expression of the cDNA segments was induced with IPTG-soaked nitrocellulose filters. After overnight incubation, the expressed proteins had been transferred to the filters and the filters were then incubated with either biotinylated SH3-domain binding peptides preconjugated to SA-AP or a monoclonal antibody recognizing the T7-Tag fusion peptide (αT7.10Mab; Novagen, Madison, WI). This antibody was used as a positive control since it recognized an epitope expressed by all the clones (part of the φ10 leader sequence common to all λEX*lox* recombinants). Sequences of pSrcI, pSrcII, Cortactin, and CaM (Calmodulin binding) peptides were Biotin-SGSGVLKRPLPIPPVTR (SEQ ID NO:42), Biotin-SGSGGILAPPVPPRNTR (SEQ ID NO:1), Biotin-SGSGSRLGEFSKPPIPQKPTWMSR (SEQ ID NO:43), and Biotin-STVPRWIEDSLRGGAARAQTRLASAK (SEQ ID NO:44), respectively.

The results are shown in Figure 16. From Figure 16 it can be seen that precomplexation with SA-AP decreases the specificity of the peptide recognition units but does not make the peptides non-specific; none of the peptides react in a significant fashion with two negative control sequences, α-actinin and calmodulin (CaM). Rather, the peptides, when precomplexed, recognize a variety of SH3 domain-containing

cDNA clones in addition to clones containing their target domains.

6.4. Characterization of cDNA clone-encoded proteins

6.4.1. Production of cDNA clone-encoded proteins

Purified DNA from all positive cDNA clones (ca. 18-20 positive clones per recognition unit) was used to transform chemical-competent BL21 cells (Hanahan et al., 1983, J. Mol. Biol. 166:557-580, the complete disclosure of which is incorporated by reference herein).

Colonies that appeared after growth overnight at 37 °C on 2xYT agar plates containing 100 µg/ml ampicillin were used to inoculate 4 ml cultures of 2xYT/amp. After 7 hours of incubation at 37 °C with shaking, IPTG was added to each culture to a final concentration of 100 µM. After an additional 2 hours of incubation, 1 ml of each culture was collected and centrifuged to pellet the cells. Cell pellets were resuspended in 400 µl 1x SDS/DTT loading buffer and boiled at 100 °C for 5 min. The resulting cell lysates were subjected to Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) on an 8% acrylamide gel. Gels were either Coomassie stained or transferred to Immobilon D membrane (Millipore) and blotted (Towbin et al., 1979, Proc. Natl. Acad. Sci. 76:4350-4354).

6.5. Materials Used in Sections 6.1, 6.2, 6.3.1, 6.3.2, 6.3.3, and 6.4.1

Blocking Solution

Hepes (pH 8)	20 mM
$MgCl_2$	5 mM
KCl	1 mM
Dithiothreitol	5 mM
Milk Powder	5% w/v

2xYT media (1L)

Bacto tryptone	16 g
Yeast Extract	10 g
NaCl	5 g

2xYT agar plates

2xYT + 15 g agar/L

2xYT top agarose (8%)
2xYT + 8 g agarose/L

SDS/DTT loading buffer
(10 mL of 5x solution)

.5 M Tris base	0.61 g
8.5% SDS	0.85 g
27.5% sucrose	2.75 g
100 mM DTT	0.154 g
.03% Bromophenol Blue	3.0 mg

Overnight cell cultures:
Inoculate media with one isolated colony of appropriate cell type and incubate 37 °C O/N with shaking

BL21 (DE3) pLysE
2xYT media

maltose	0.2%
$MgSO_4$	10 mM
Chloramphenicol	25 µg/mL

BM25.8
2xYT media

maltose	0.2%
$MgSO_4$	10 mM
Chloramphenicol	34 µg/ml
Kanamycin	50 µg/ml

6.6. Other Functional Domains and Recognition Units

In a manner similar to that described above for SH3 domains, recognition units directed to other functional domains of interest can be chosen for use in the present method. For example, as recognition units for a study of GST functional domains, the following GST-binding peptides can be used to screen a plurality of polypeptides: Class I CWSEWDGNEC (SEQ ID NO:46), CGQWADDGYC (SEQ ID NO:47), CEOWDGYGAC (SEQ ID NO:48), CWPFWDGSTC (SEQ ID NO:49), CMIWPDGEEC (SEQ ID NO:50), CESOWDGYDC (SEQ ID NO:51), CQQWKEDGWC (SEQ ID NO:52), or CLYOWDGYEC (SEQ ID NO:53); Class II - CMGDNLGDDC (SEQ ID NO:54), CMGDSLGOSC (SEQ ID NO:55), CMDDDLGKGC (SEQ ID NO:56), CMGENLGWSC (SEQ ID NO:57), or CLGESLGWMC (SEQ ID NO:58).

Moreover, the following SH2-binding peptides can be used according to the methods of the present invention to

identify SH2 domain-containing polypeptides: GDGYEEISP (SEQ ID NO:59) (for Src family), GDGYDEPSP (SEQ ID NO:60) (for Nck), GDGYDHPSP (SEQ ID NO:61) (for Crk), GDGYVIPSP (SEQ ID NO:62) (PLCγN), GDGYQNYSP (SEQ ID NO:63) (for PLCγC), GDGYMAMSP (SEQ ID NO:64) (for p85PI3KN and p85PI3KC), or GDGQNYSP (SEQ ID NO:65) (for Grb2). See, Yang, Cell 72:767-778, the complete disclosure of which is incorporated by reference herein.

Further, polypeptides with a "PH" functional domain (analogous to the proteins Vav, Bcr, Msos, PLCδ, Atk, or Pleckstrin) can be identified using PH-binding peptides, such as those described by Mayer et al., Cell 73:629-630, the complete disclosure of which is incorporated by reference herein.

Other recognition units can be readily contemplated, including other synthetic, semisynthetic, or naturally derived molecules.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

WHAT IS CLAIMED IS:

1. A method of identifying a polypeptide comprising a functional domain of interest comprising:

(a) contacting a multivalent recognition unit complex with a plurality of polypeptides; and

(b) identifying a polypeptide having a selective binding affinity for said recognition unit complex.

2. The method of claim 1 in which said plurality of polypeptides is from a polypeptide expression library.

3. The method of claim 1 in which said plurality of polypeptides is obtained from a virus.

4. The method of claim 2 in which said expression library is a cDNA expression library.

5. The method of claim 2 in which said expression library is a genomic DNA library.

6. The method of claim 2 in which said expression library is a recombinant bacteriophage library.

7. The method of claim 6 in which said recombinant bacteriophage library is a recombinant M13 library.

8. The method of claim 2 in which said expression library is a recombinant plasmid or cosmid library.

9. The method of claim 1 in which the recognition unit is a peptide.

10. The method of claim 1 in which said recognition unit is a peptide having less than about 140 amino acid residues.

11. The method of claim 1 in which said recognition unit is a peptide having less than about 100 amino acid residues.

12. The method of claim 1 in which said recognition unit is a peptide having less than about 70 amino acid residues.

13. The method of claim 1 in which said recognition unit is a peptide having about 6 to 60 amino acid residues.

14. The method of claim 1 in which said recognition unit is a peptide having 20 to 50 amino acid residues.

15. The method of claim 1 in which the valency of the recognition unit in the complex is at least two.

16. The method of claim 9 in which the valency of the recognition unit in the complex is at least two.

17. The method of claim 1 in which the valency of the recognition unit in the complex is at least four.

18. The method of claim 9 in which the valency of the recognition unit in the complex is at least four.

19. The method of claim 17 in which the recognition unit complex is a complex comprising (a) avidin or streptavidin, and (b) biotinylated recognition units.

20. The method of claim 18 in which the recognition unit complex is a complex comprising (a) avidin or streptavidin, and (b) the biotinylated peptides.

21. The method of claim 2 in which said identifying step comprises selecting a positive clone, which harbors a DNA construct encoding a polypeptide having a selective affinity for said recognition unit and which polypeptide includes the functional domain of interest or a functional equivalent thereof.

22. The method of claim 21 which further comprises determining the coding sequence of said DNA construct.

23. The method of claim 22 which further comprises deducing an amino acid sequence from said coding sequence.

24. The method of claim 1 in which said contacting step comprises immobilizing said recognition unit complex on a solid support and bringing a solution containing said plurality of polypeptides in contact with said immobilized recognition unit complex.

25. The method of claim 1 in which said contacting step comprises separating said plurality of polypeptides and bringing a solution of said recognition unit complex in contact with said separated polypeptides.

26. The method of claim 1 in which said identifying step includes selecting a polypeptide, among said plurality of polypeptides, having a selective affinity for said recognition unit and determining the amino acid sequence of said polypeptide.

27. The method of claim 1 in which said plurality of polypeptides is immobilized on a solid support.

28. The method of claim 27 in which said contacting step comprises contacting said solid support with a solution containing said recognition unit complex.

29. The method of claim 28 which further comprises washing away any unbound recognition unit complex.

30. The method of claim 29 which further comprises detecting any recognition unit complex that remains bound to said solid support.

31. The method of claim 1 in which said selective binding affinity is on the order of about 1 nM to about 1 mM.

32. The method of claim 1 in which said selective binding affinity is on the order of about 10 nM to about 100 μM.

33. The method of claim 1 in which said selective binding affinity is on the order of about 100 nm to about 10 μM.

34. The method of claim 1 in which said selective binding affinity is on the order of about 100 nm to about 1 μM.

35. The method of claim 9 in which said peptide is chosen from a random peptide library.

36. A method of identifying a polypeptide comprising a functional domain of interest comprising:

(a) contacting a multivalent recognition unit complex, which complex comprises (i) avidin or streptavidin, and (ii) biotinylated recognition units, with a plurality of polypeptides from a cDNA expression library, in which the recognition unit is a peptide having in the range of 6 to 60 amino acid residues; and

(b) identifying a polypeptide having a selective binding affinity for said recognition unit complex.

37. The method of claim 4 or 36 in which the cDNA expression library is a human cDNA expression library.

38. The method of claim 36 in which the peptide is previously identified by a method comprising screening a random peptide library to identify a peptide having selective binding affinity for the functional domain of interest or a functional equivalent thereof.

39. The method of claim 36 in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zippers, and helix-turn-helix.

40. The method of claim 1 in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix.

41. The method of claim 1, 37, or 38 in which the functional domain of interest is an SH3 domain.

42. A method of identifying a polypeptide comprising an SH3 domain of interest comprising:

(a) contacting a multivalent recognition unit complex, which complex comprises (i) avidin or streptavidin, and (ii) biotinylated recognition units, with a plurality of polypeptides from a cDNA expression library, in which the recognition unit is a peptide having in the range of 6 to 60 amino acid residues and which selectively binds an SH3 domain; and

(b) identifying a polypeptide having a selective binding affinity for said recognition unit complex.

43. The method of claim 1 in which the functional domain of interest comprises a catalytic site.

44. The method of claim 43 in which said catalytic site corresponds to that found in glutathione S-transferase.

45. A method of identifying a polypeptide comprising a functional domain of interest or a functional equivalent thereof comprising:

(a) screening a random peptide library to identify a peptide that selectively binds a functional domain of interest; and

(b) screening a cDNA or genomic expression library with said peptide or a binding portion thereof to identify a polypeptide that selectively binds said peptide.

46. The method of claim 45 in which the screening step (b) is carried out by use of said peptide in a multivalent peptide complex.

47. The method of claim 46 in which the screening step (b) is carried out by use of said peptide in a complex comprising streptavidin and biotinylated peptide.

48. The method of claim 46 in which the screening step (b) is carried out by use of said peptide in the form of multiple antigen peptides (MAP).

49. The method of claim 46 in which the screening step (b) is carried out by use of said peptide cross-linked to bovine serum albumin or keyhole limpet hemocyanin.

50. A method of identifying a polypeptide comprising a functional domain of interest or a functional equivalent thereof comprising:

(a) screening a random peptide library to identify a plurality of peptides that selectively bind a functional domain of interest;

(b) determining at least part of the amino acid sequences of said peptides;

(c) determining a consensus sequence based upon the determined amino acid sequences of said peptides; and

(d) screening a cDNA or genomic expression library with a peptide comprising the consensus sequence to identify a polypeptide that selectively binds said peptide.

51. The method of claim 50 in which the screening step (d) is carried out by use of said peptide in a multivalent peptide complex.

52. A method of identifying a polypeptide comprising a functional domain of interest or a functional equivalent thereof comprising:

(a) screening a random peptide library to identify a first peptide that selectively binds a functional domain of interest;

(b) determining at least part of the amino acid sequence of said first peptide;

(c) searching a database containing the amino acid sequences of a plurality of expressed natural proteins to identify a protein containing an amino acid sequence homologous to the amino acid sequence of said first peptide; and

(d) screening a cDNA or genomic expression library with a second peptide comprising the sequence of said protein that is homologous to the amino acid sequence of said first peptide.

53. An assay kit comprising in one or more containers:

(a) a purified polypeptide containing a functional domain of interest, in which the functional domain of is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix; and

(b) a purified recognition unit having a selective binding affinity for said functional domain in said polypeptide.

54. The assay kit of claim 53 in which said polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

55. The assay kit of claim 53 in which said polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:113-115, 118-121, 125-128, 133-139, 204-218, and 219.

56. The assay kit of claim 53 in which said recognition unit is a peptide.

57. The assay kit of claim 53 in which said polypeptide or recognition unit is labeled.

58. The assay kit of claim 57 in which said polypeptide or recognition unit is labeled with an enzyme.

59. The assay kit of claim 57 in which said polypeptide or recognition unit is labeled with an epitope.

60. The assay kit of claim 57 in which said polypeptide or recognition unit is labeled with a chromogen.

61. The assay kit of claim 57 in which said polypeptide or recognition unit is labeled with biotin.

62. The assay kit of claim 53 in which said polypeptide or recognition unit is immobilized on a solid support.

63. An assay kit comprising in containers:

(a) a plurality of purified polypeptides, each polypeptide in a separate container and each polypeptide containing a functional domain of interest in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix; and

(b) at least one recognition unit having a selective binding affinity for said functional domain in each of said plurality of polypeptides.

64. An assay kit comprising in one or more containers:

(a) a plurality of purified polypeptides, each polypeptide in a separate container and each polypeptide containing an SH3 domain; and

(b) at least one peptide having a selective affinity for the SH3 domain in each of said plurality of polypeptides.

65. A kit comprising a plurality of purified polypeptides comprising a functional domain of interest, each polypeptide in a separate container, and each polypeptide having a functional domain of a different sequence but capable of displaying the same binding specificity.

66. The kit of claim 65 in which the polypeptides have an amino acid sequence selected from the group consisting of: SEQ ID NO: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

67. The kit of claim 65 in which the functional domain is an SH3 domain.

68. The kit of claim 65 in which the functional domain is an SH3 domain from a polypeptide having an amino acid sequence selected from the group consisting of: SEQ ID NO: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

69. A method for screening a potential drug candidate comprising:

(a) allowing at least one polypeptide comprising a functional domain of interest to come into contact with at least one recognition unit having a selective affinity for said functional domain in said polypeptide, in the presence of an amount of a potential drug candidate, such that said polypeptide and said recognition unit are capable of interacting when brought into contact with one another in the

absence of said drug candidate, and in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix; and

(b) determining the effect, if any, of the presence of the amount of said drug candidate on the interaction of said polypeptide with said recognition unit.

70. The method of claim 69 in which the effect of the drug candidate upon multiple, different interacting polypeptide-recognition unit pairs is determined in which at least some of said polypeptides have a functional domain that differs in sequence but is capable of displaying the same binding specificity as the functional domain in another of said polypeptides.

71. The method of claim 69 in which at least one of said at least one polypeptide or recognition unit contains a consensus functional domain and consensus recognition unit, respectively.

72. The method of claim 69 in which the polypeptide is a polypeptide identified by the method of claim 1.

73. The method of claim 69 in which the drug candidate is an inhibitor of the polypeptide-recognition unit interaction that is identified by detecting a decrease in the binding of polypeptide to recognition unit in the presence of such inhibitor.

74. A purified polypeptide comprising an SH3 domain, said SH3 domain having an amino acid sequence selected from the group consisting of: SEQ ID NOs:113-115, 118-121, 125-128, 133-139, 204-218, and 219.

75. A purified polypeptide comprising an SH3 domain, said polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

76. A purified DNA encoding an SH3 domain, said DNA having a sequence selected from the group consisting of SEQ ID NOs: 7, 9, 11, 17, 19, 21, 23, 29, 31, 37, 39, 189, 191, 193, 195, 197, 199, and 220.

77. A purified DNA encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

78. A purified DNA encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: SEQ ID NOs:113-115, 118-121, 125-128, 133-139, 204-218, and 219.

79. A purified molecule comprising an SH3 domain of a polypeptide having an amino acid sequence selected from the group consisting of: SEQ ID NO: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

80. A fusion protein comprising (a) an amino acid sequence comprising an SH3 domain of a polypeptide having the amino acid sequence of SEQ ID NO: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, or 221 joined via a peptide bond to (b) an amino acid sequence of at least six amino acids from a different polypeptide.

81. A purified DNA encoding the fusion protein of claim 80.

82. A nucleic acid vector comprising the DNA of claim 81.

83. A nucleic acid vector comprising the DNA of claim 76.

84. A nucleic acid vector comprising the DNA of claim 78.

85. A recombinant cell containing the nucleic acid vector of claim 82, 83, or 84.

86. A purified nucleic acid hybridizable to a nucleic acid having a sequence selected from the group consisting of: SEQ ID NOs: 7, 9, 11, 17, 19, 21, 23, 29, 31, 37, 39, 189, 191, 193, 195, 197, 199, and 220.

87. A method of producing the fusion protein of claim 80 comprising culturing a recombinant cell containing a nucleic acid vector encoding said fusion protein such that said fusion protein is expressed, and recovering the expressed fusion protein.

88. A method of producing the polypeptide of claim 74 comprising culturing a recombinant cell containing a nucleic acid vector encoding said polypeptide such that said polypeptide is expressed, and recovering the expressed polypeptide.

89. The method of claim 69 in which said polypeptide is a polypeptide containing an SH3 domain produced by a method comprising:

(i) screening a peptide library with an SH3 domain to obtain one or more peptides that bind the SH3 domain;

(ii) using one of the peptides from step (i) to screen a source of polypeptides to identify one or more polypeptides containing an SH3 domain;

(iii) determining the amino acid sequence of the polypeptides identified in step (ii); and

(iv) producing the one or more novel polypeptides containing an SH3 domain.

90. The method of claim 69 in which said polypeptide is a polypeptide containing an SH3 domain produced by a method comprising:

(i) screening a peptide library with an SH3 domain to obtain a plurality of peptides that bind the SH3 domain;

(ii) determining a consensus sequence for the peptides obtained in step (i);

(iii) producing a peptide comprising the consensus sequence;

(iv) using the peptide comprising the consensus sequence to screen a source of polypeptides to identify one or more polypeptides containing an SH3 domain;

(v) determining the amino acid sequence of the polypeptides identified in step (iv); and

(vi) producing the one or more polypeptides containing an SH3 domain.

91. A method of determining the potential pharmacological activities of a molecule comprising:

(a) contacting the molecule with a compound comprising a functional domain under conditions conducive to binding;

(b) detecting or measuring any specific binding that occurs; and

(c) repeating steps (a) and (b) with a plurality of different compounds, each compound comprising a functional domain of different sequence but capable of displaying the same binding specificity.

92. The method of claim 91 in which the functional domain is an SH3 domain.

93. The method of claim 92 in which the compounds comprise the SH3 domains of Src, Abl, Cortactin, Phospholipase

Cγ, Nck, Crk, p53bp2, Amphiphysin, Grb2, RasGap, or Phosphatidylinositol 3' kinase.

94. A method of identifying a compound that affects the binding of a molecule comprising a functional domain to a recognition unit that selectively binds to the functional domain comprising:

(a) contacting the molecule comprising the functional domain and the recognition unit under conditions conducive to binding in the presence of a candidate compound and measuring the amount of binding between the molecule and the recognition unit and in which the functional domain of interest is a domain selected from the group consisting of an SH1, SH2, SH3, PH, PTB, LIM, armadillo, Notch/ankyrin repeat, zinc finger, leucine zipper, and helix-turn-helix;

(b) comparing the amount of binding in step (a) with the amount of binding known or determined to occur between the molecule and the recognition unit in the absence of the candidate compound, where a difference in the amount of binding between step (a) and the amount of binding known or determined to occur between the molecule and the recognition unit in the absence of the candidate compound indicates that the candidate compound is a compound that affects the binding of the molecule comprising a functional domain and the recognition unit.

95. The method of claim 94 in which the functional domain is an SH3 domain.

96. The method of claim 20 in which the recognition unit complex is a complex comprising (a) streptavidin conjugated to alkaline phosphatase; and (b) the biotinylated peptides.

97. A method of identifying a polypeptide comprising a functional domain of interest comprising:

(a) contacting a recognition unit that is a peptide having 140 amino acids or fewer with a plurality of polypeptides; and

(b) identifying a polypeptide having a selective binding affinity for said recognition unit complex.

98. An antibody to a polypeptide comprising an amino acid sequence selected from the group consisting of: SEQ ID NOs:113-115, 118-121, 125-128, 133-139, 204-218, and 219.

99. An antibody to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 8, 10, 12, 18, 20, 22, 24, 30, 32, 38, 40, 190, 192, 194, 196, 198, 200, and 221.

100. The purified nucleic acid of claim 86 that is a human nucleic acid encoding a polypeptide containing a functional domain.

101. A purified protein encoded by a first nucleic acid comprising a human cDNA or genomic sequence hybridizable to a second nucleic acid having a sequence selected from the group consisting of: SEQ ID NOs:7, 9, 11, 17, 19, 21, 29, and 31.

102. The assay kit of claim 53 in which said polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:6, 14, 16, 26, 28, 34, 36, 112, 116, 117, 122-124, 129-132, and 140.

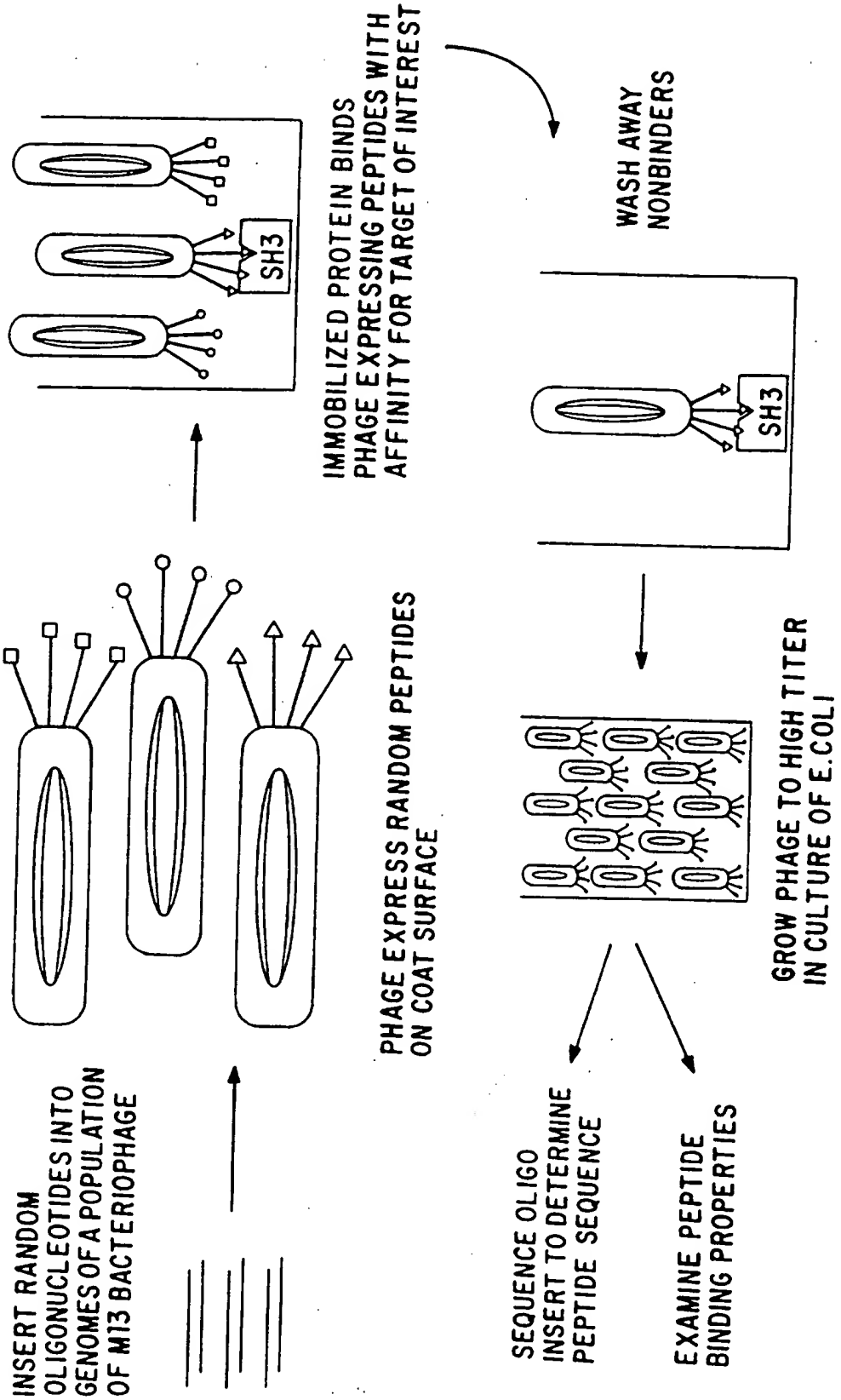
INSERT RANDOM OLIGONUCLEOTIDES INTO GENOMES OF A POPULATION OF M13 BACTERIOPHAGE
PHAGE EXPRESS RANDOM PEPTIDES ON COAT SURFACE
IMMOBILIZED PROTEIN BINDS PHAGE EXPRESSING PEPTIDES WITH AFFINITY FOR TARGET OF INTEREST
SH3
WASH AWAY NONBINDERS
SH3
GROW PHAGE TO HIGH TITER IN CULTURE OF E.COLI
SEQUENCE OLIGO INSERT TO DETERMINE PEPTIDE SEQUENCE
EXAMINE PEPTIDE BINDING PROPERTIES

FIG. 1

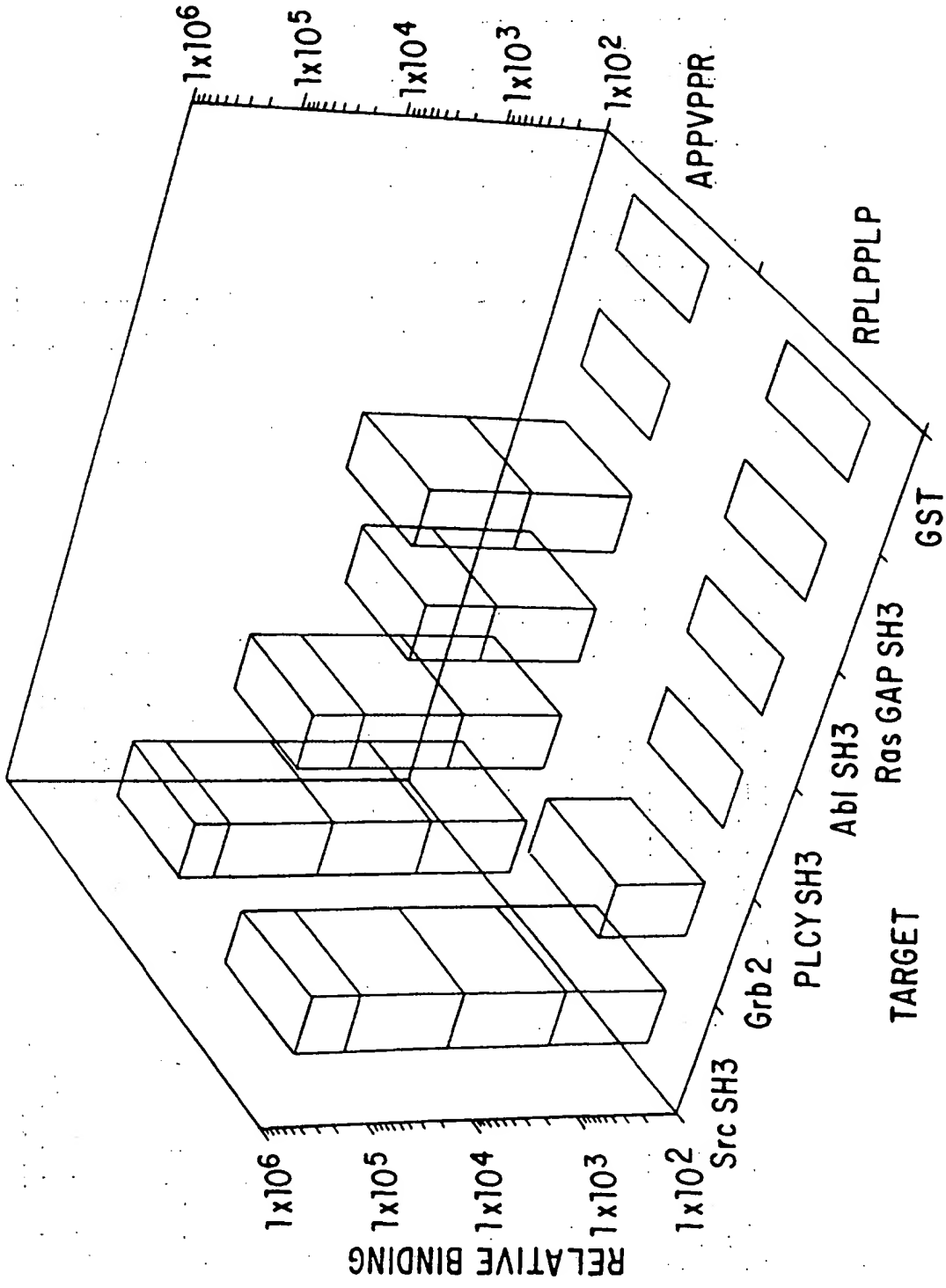
RELATIVE BINDING
1x10^6
1x10^5
1x10^4
1x10^3
1x10^2
Src SH3
Grb 2
PLCY SH3
Abl SH3
Ras GAP SH3
GST
TARGET
APPVPPR
RPLPPLP

FIG. 2

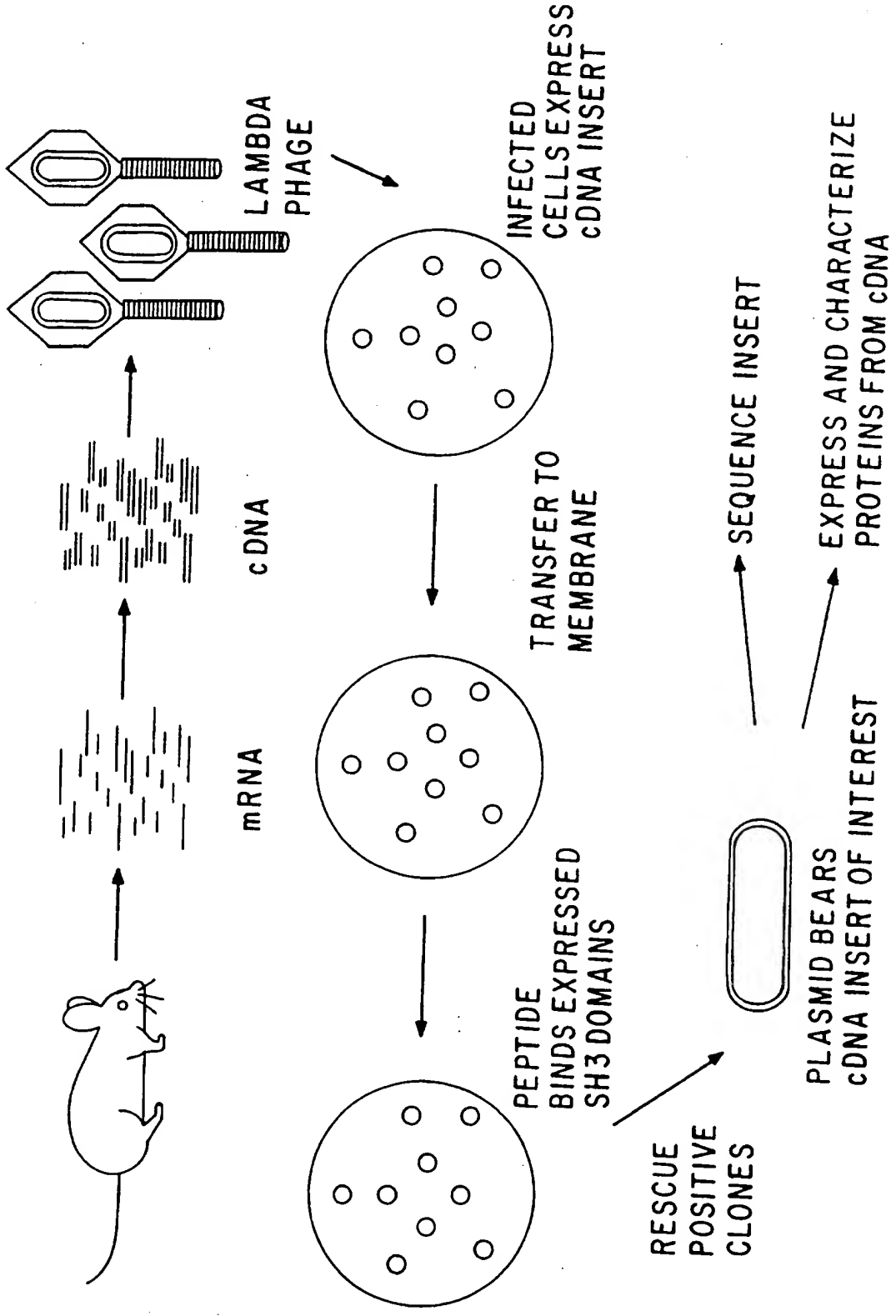
LAMBDA PHAGE
INFECTED CELLS EXPRESS cDNA INSERT
cDNA
TRANSFER TO MEMBRANE
SEQUENCE INSERT
EXPRESS AND CHARACTERIZE PROTEINS FROM cDNA
mRNA
PEPTIDE BINDS EXPRESSED SH3 DOMAINS
PLASMID BEARS cDNA INSERT OF INTEREST
RESCUE POSITIVE CLONES

FIG. 3

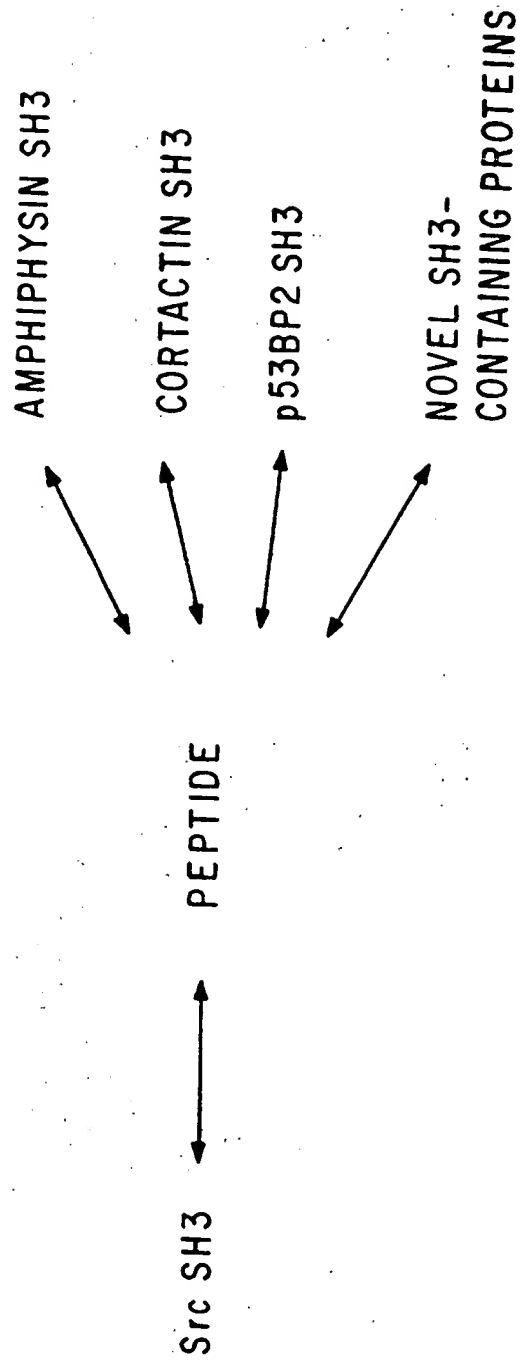
AMPHIPHYSIN SH3
CORTACTIN SH3
p53BP2 SH3
NOVEL SH3-CONTAINING PROTEINS
PEPTIDE
Src SH3

FIG. 4

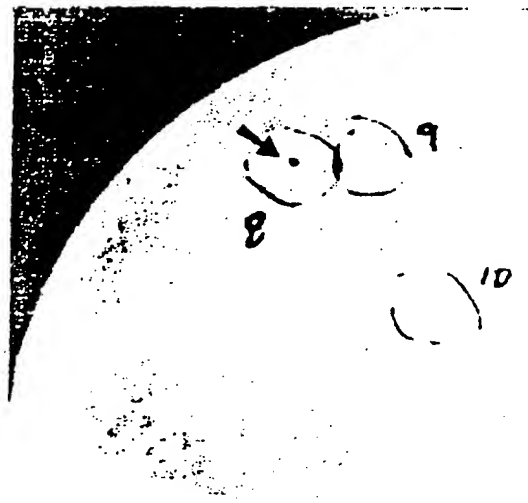

FIG.5A

FIG.5B

```
Sc_Fus1_Sh3: TVI QDYE PRLTDE..... .....IRISLG EKVKIL...A .....THTD  CLVEKCNTRK
Sc_Bob1_Sh3: RAL FDYD KTKDCGFLSQ ...ALSFRFG DVLHVIDA.....GDEE  QAR RVHSDSE
Hs_Mpp1_Sh3: RAQ FDYD PKKDNLIPCK .EAGLKFATG DIIQIINK.....DDSN  QGR ..VEGSS
  Hs_Zo1_Sh3: RTH PHYE KESPY..... ....GLSFNKG EVFRAVDTLY...NGKLGS AI RIGKNHKE
Hs_Ncf1_Csh3: VAI KAYT AV.EGD..E. .....VSLLEG EAVEVIHK.....LLDG  V IRKDDVTGY
Hs_Rasgap_Sh3: RAI LPYT KVPDTD..E. .....ISFLKG DMFIVHNE.....LED  ..MWVTNLRT
  Sc_Sla_Nsh3: RAV YAYE ..PQTP..EE ....LAIQED DLLYLLQKSD.....IDD  TVKKR VIGSD
Sc_Bem1_Nsh3: KAK YSYQ ..AQTS..KE ....LSFMEG EFFYV..SGD.....EKD  KASNP STGKE
Hs_Ncf2_Nsh3: AHR VLFG FVPETK..EE ....LQVMPG NIVFVL..KK .....GND  ATVMFNG.QK
  Sc_Sla_Csh3: RGI VQYD FMAESQ..DE .....LTIKSG DKVYILDDKK .....SKD  MCQLVDS.GK
Hs_Grb2_Csh3: QAL FDPD ..PQED..GE ....LGFRRG DFIHVMD..N S.....DPN  KG.A.CH.GQ
  Hs_Nck_Csh3: QAL YPFS ..SSND..EE ....LNFEKG DVMDVIEKPE N.....DPE  KCRK.IN.GM
   Mm_Tec_Sh3: VAM YDFG ..ATEA..HD ....LRLERG QEYIILEKN.  .....DLH  RARD.KN.GQ
   Hs_Atk_Sh3: VAL YDYN ..PMNA..ND ....LQLRKG DEYFILEES.  .....NLP  RARD.KY.GW
   Hs_Abl_Sh3: VAL YDFV ..ASGD..NT ....LSITKG EKLRVLGYN.  .....HNGE  EAQT.KN.GQ
   Hs_Src_Sh3: VAL YDYE ..SRTE..TD ....LSFKKG ERLQIVNNT.  .....EGD  LAHSLST.GQ
   Hs_Fgr_Sh3: IAL YDYE ..ARTE..DD ....LTFTKG EKFHILNNT.  .....EGD  EARSLSS.GK
   Hs_Fyn_Sh3: VAL YDYE ..ARTE..DD ....LSFHGK EKFQILNSS.  .....EGD  EARSLTT.GE
  Hs-Yes_Sh3: VAL YDYE ..ARTT..ED ....LSFKKG ERFQIINNT.  .....EGD  EARSIAT.GK
   Mm_Fgr_Sh3: VAL YDYE ..ARTG..DD ....LTFTKG EKFHILNNT.  .....EYD  EARSLSS.GH
   Hv_Stk_Sh3: VAL YDYE ..ARIS..ED ....LSFKKG ERLQIINTA.  .....DGD  YARSLIT.NS
   Hs_Hck_Sh3: VAL YDYE ..AIHH..ED ....LSFQKG DQMVVLEES.  .....GE  KARSLAT.RK
   Hs_Lyn_Sh3: VAL YPYD ..GIHP..DD ....LSFKKG EKMKVLEEH.  .....GE  KAKSLLT.KK
   Mm_Blk_Sh3: VAL PDYA ..AVND..RD ....LQVLKG EKLQVLRST.  .....GD  LARSLVT.GR
   Hs_Lck_Sh3: IAL HSYE ..PSHD..GD ....LGFEKG EQLRILEQS.  .....GE  KAQSLTT.GQ
Hs_Nck_Nsh3: VAK FDYV ..AQQE..QE ....LDIKKN ERLWLLDDSK SW.....  VRNSM...NK
Sc_Sla_Msh3: RAI YDYE QVQNAD..EE ....LTFHEN DVFDVFKKD. .....DAD  LVKSTVS.NE
```

FIG. 6A

```
Hs_Hs1_Sh3:   VAL YDYQ ..GEGS..DE .....LSFDPD DVITDIEMV. .....DEG  RG..RCH.GH
Sc_Abp1_Sh3: TAE YDYD ..AAED..NE .....LTFVEN DKIINIEFV. .....DDD  LGELEKD.GS
Hs_Nck_Msh3: YVK FNYN ..AERE..DE .....LSLIKG TKVIVMEKC. .....SDG  RG..SYN.GQ
Hs_Vav_Sh3: KAR YDFC ..ARDR..SE .....LSLKEG DIIKILNKK. .....GQQ  WRGEIY..GR
Hs_Grb2_Nsh3: IAK YDFK ..ATAD..DE .....LSFKRG DILKVLNEE. .....CDQ  YKAELN..GK
Hs_Plcg2_Sh3: KAL YDYK ..AKRS..DE .....LSFCRG ALIHNVSKE .....PG  WKGDYGT.RI
Hs_Plcg1_Sh3: KAL FDYK ..AQRE..DE .....LTFIKS AIIQNVEKQ. .....EG  WRGDYGG.KK
Ac_Mys1b_Sh3: KAL YDYD ..AQTG..DE .....LTFKEG DTIIVHQKD. .....PA  WEGELN..GK
Ac_Mys1c_Sh3: RAL YDFA ..AENP..DE .....LTFNEG AVVTVINKS. .....NP  WEGELN..GQ
Dd_Mys1b_Sh3: KAL YDYD ..ASST..DE .....LSFKEG DIIFIVQKD. .....NG  TQGELKS.GQ
Hs_Ncf2_Csh3: EAL FSYE ..ATQP..ED .....LEFQEG DIILVLSKV. .....NE  LEGECK..GK
Hs_Ncf1_Nsh3: RAI ADYE ..KTSG..SE .....MALSTG DVVEVVEKS. .....ESG  FCQM..K.AK
Hs_Spectrin_Sh3: MAL VDFQ ..ARSP..RE .....VTMKKG DVLTLLSSI. .....NKD  KVEA..A.DH
Sc_Bem1_Csh3: YAI VLYD FKAEKA..DE .....LTTYVG ENLFICAH.. .....HNCE  IAKPIGRLGG
Sc_Cde25_Sh3: VAA YDFN YPIKKD..SS .SQLLSVQQG ETIYILNKNS      SG  ...DGL VIDD
Sp_Ste_Sh3:   .MR FQTT AISDYENSSN .PSFLKFSAG DTIIVIEVLE       D  CDG.....
Hs_Pf3ka_Sh3: RAL VDYK KEREEDIDLH LGDILTVNKG SLVALGFSDG QEARPEEI  LNGYNETTGE
```

FIG. 6B

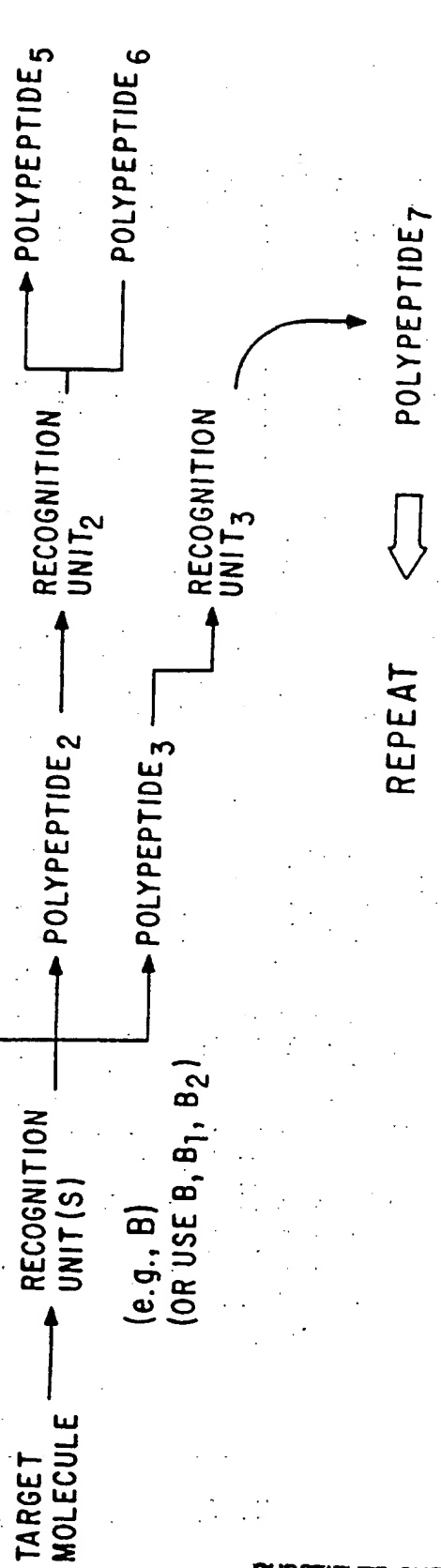
TARGET MOLECULE
RECOGNITION UNIT (S)
(e.g., B)
(OR USE B, B1, B2)
POLYPEPTIDE1
POLYPEPTIDE2
POLYPEPTIDE3
RECOGNITION UNIT1
RECOGNITION UNIT2
RECOGNITION UNIT3
(e.g., B2)
POLYPEPTIDE4
POLYPEPTIDE5
POLYPEPTIDE6
POLYPEPTIDE7
REPEAT

FIG. 7B

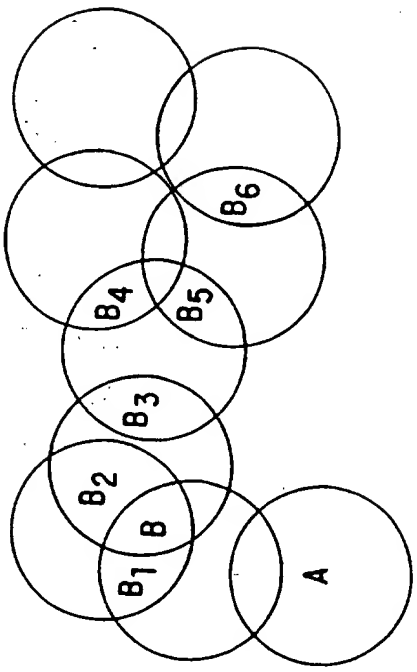
A
B1
B
B2
B3
B4
B5
B6

FIG. 7A

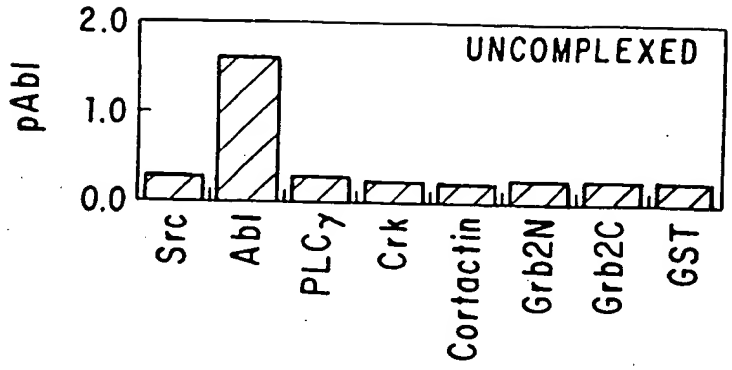

FIG. 8A

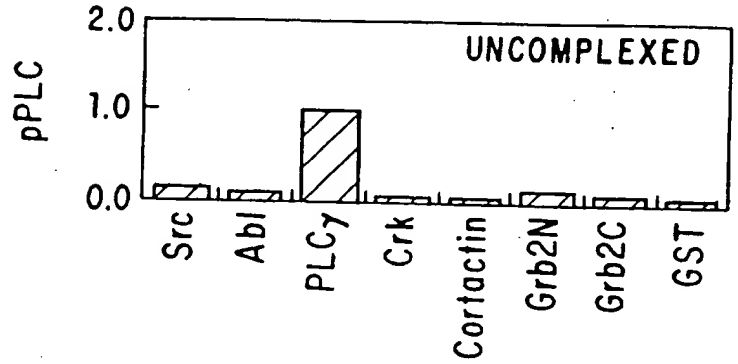

FIG. 8B

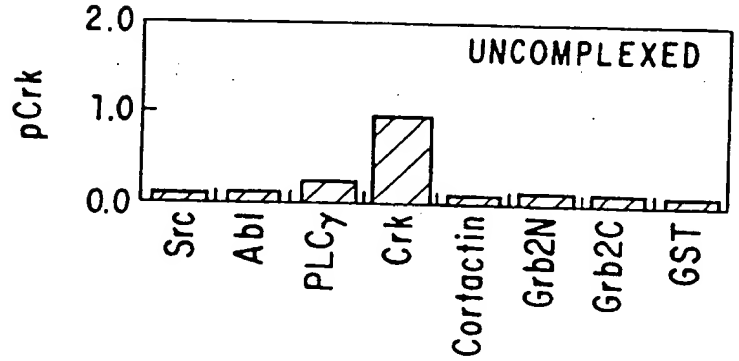

FIG. 8C

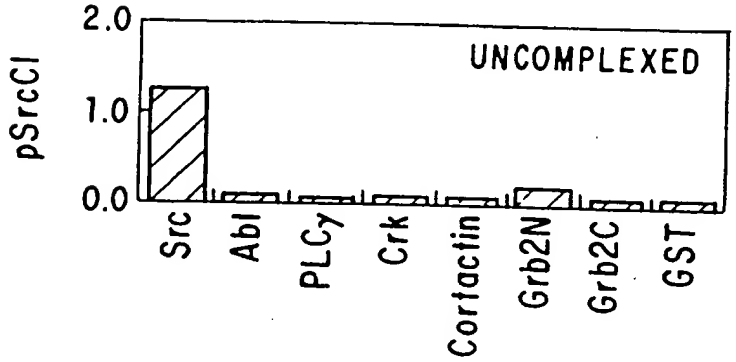

FIG. 8D

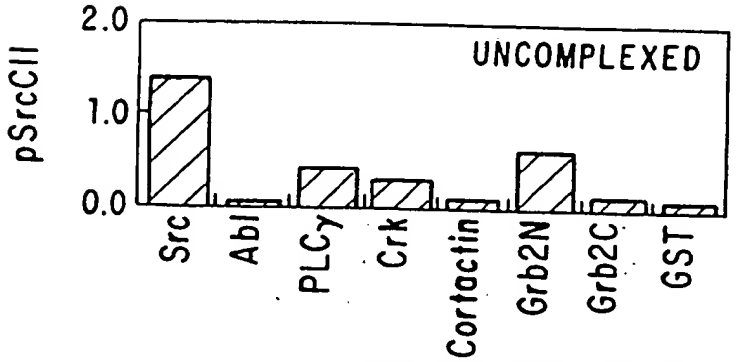

FIG. 8E

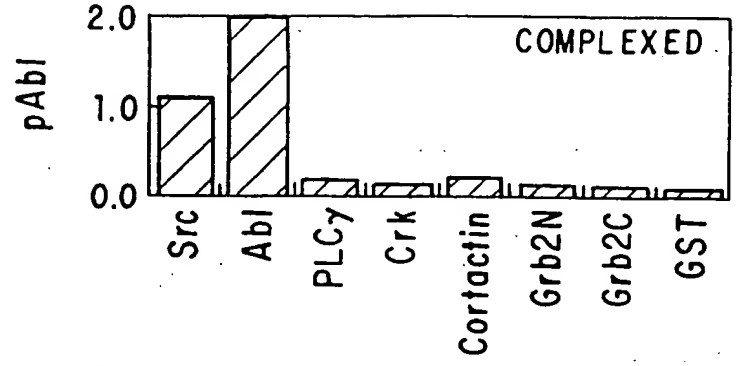

FIG. 8F

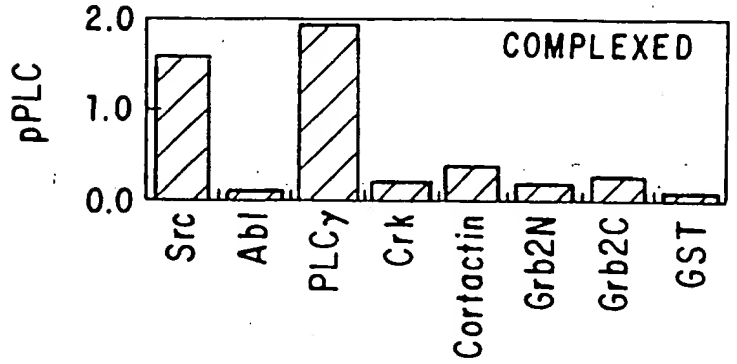

FIG. 8G

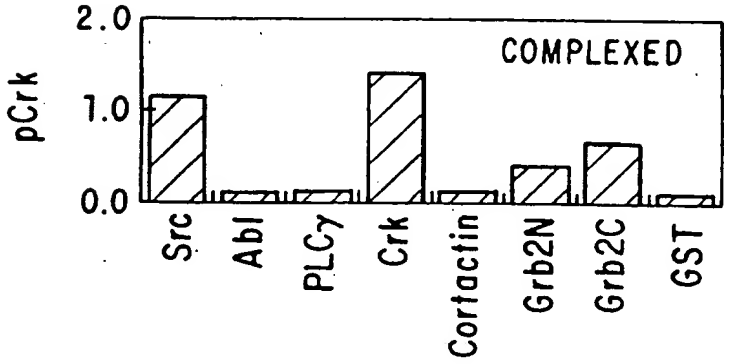

FIG. 8H

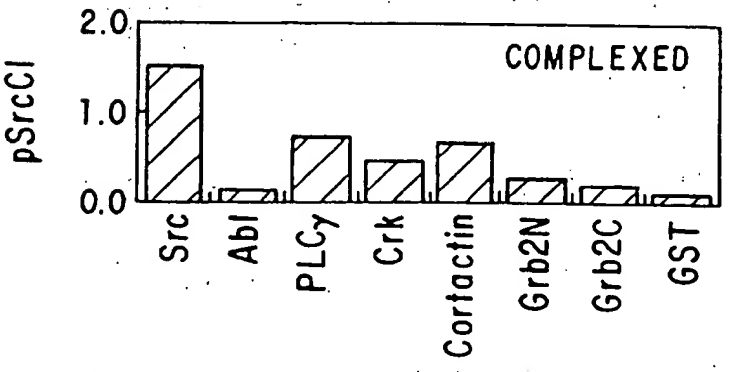

FIG. 8I

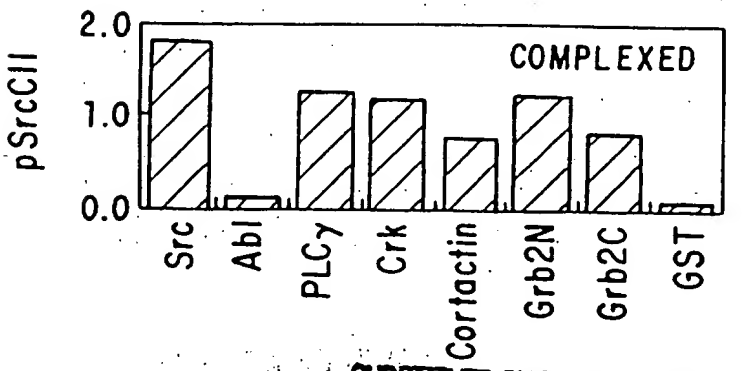

FIG. 8J

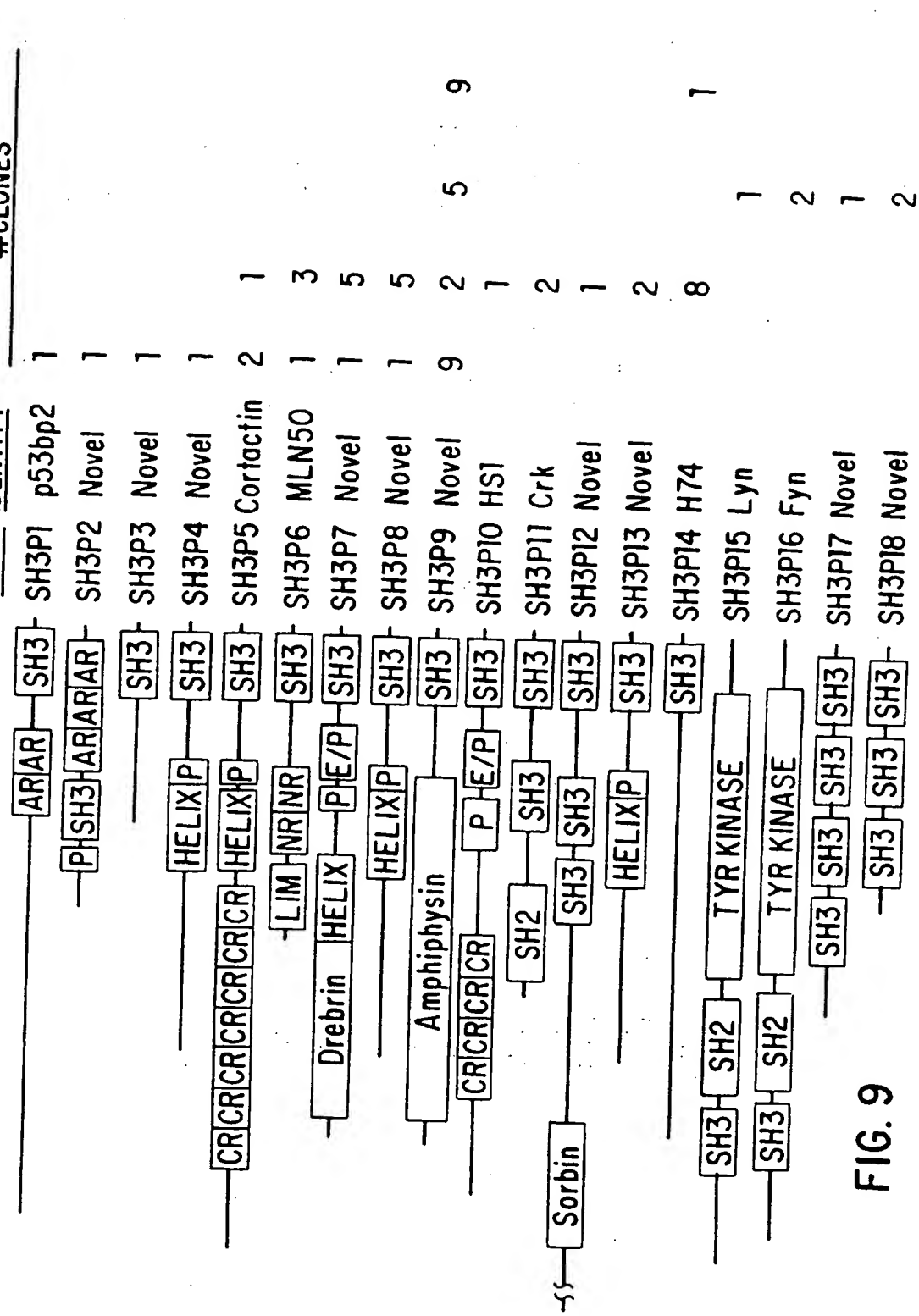
LIGAND: pSrcCII pCort pSrcCII pSrcCII
LIBRARY: Mouse Mouse Human1 Human 2
NAME IDENTITY #CLONES
SH3P1 p53bp2 1
SH3P2 Novel 1
SH3P3 Novel 1
SH3P4 Novel 1
SH3P5 Cortactin 2 1
SH3P6 MLN50 1 3
SH3P7 Novel 1 5
SH3P8 Novel 1 5
SH3P9 Novel 9 2 5 9
SH3P10 HS1 1
SH3P11 Crk 2
SH3P12 Novel 1
SH3P13 Novel 2
SH3P14 H74 8 1
SH3P15 Lyn 1
SH3P16 Fyn 2
SH3P17 Novel 1
SH3P18 Novel 2
SH3
AR
P
HELIX
CR
LIM
NR
E/P
Drebrin
Amphiphysin
SH2
Sorbin
TYR KINASE

FIG. 9

```
SH3P1   p53bp2     NKGTVYALWDYEAQNSDELSFHEGDAITILRRKDEN..ETEWWWARLG.........DREGYVPKNLLGLY
SH3P2   Novel      QVKVFRALYTFEPRTPDELYFEEGDIIYITDMS.....DTSWWKGTCK.........GRTGLIPSNYVAEQ
SH3P3   Novel      HWTPYRAMYQYRPQNEDELELREGDRVDVMQQC.....DDGWFVGVSRRT.......QKFGTFPGNYVAPV
SH3P4   Novel      DQPSCKALYDFEPENDGELGFREGDLITLTNQI.....DENWYEGMLH.........GQSGFFPLSYVQVL
SH3P5   Cortactin LGITAIALYDYQAAGDDEISFDPDDIITNIEMI.....DDGWWRGVCK.........GRYGLFPANYVELR
SH3P6   MLN50      GGKRYRAVYDYSAADEDEVSFQDGDTIVNVQQI.....DDGWMYGTVERT.......GDTGMLPANYVEAI
SH3P7   Novel      QGLCARALYDYQAADDTEISFDPENLITGIEVI.....DEGWWRGYGPD.........GHFGMFPANYVELI
SH3P8   Novel      DQPCCRALYDLEPENEGELAFKEGDIITLTNQI.....DENWYEGMLH.........GQSGFFPINYVEIL
SH3P9   Novel, m   FMFKVQAQHDYTATDTDELQLKAGDVVLVIPFQNPEEQDEGWLMGVKESDWNQHKELEKCRGVFPENFTERV
        Novel, h   FMFKVQAQHDYTATDTDELQLKAGDVVLVIPFQNPEEQDEGWLMGVKESDWNQHKELEKCRGVFPENFTERV
SH3P10  HS1        AGISAIALYDYQGEGSDELSFDPDDIITDIEMV.....DEGWWRGQCR.........GHFGLFPANYVKLL
SH3P11  Crk A      EAEYVRALFDFNGNDEEDLPFKKGDILRIRDKP.....EEQWWNAEDSE.......GKRGMIPVPYVEKY
            B      RVIQKRVPNAYDKTALALEVGELVKVTKINV.......SGQWEGECN.........GKRGHFPFTHVRLL
SH3P12  Novel A    EMRPARAKFDFKAQTLKELPLQKGDVVYIYRQI.....DQNWYEGEHH.........GRVGIFPRTYIELL
              B    EYGEAIAKFNFNGDTQVEMSFRKGERITLLRQV.....DENWYEGRIPGT.......SRQGIFPITYVDVL
              C    DLCSYQALYSYVPQNDDELELRDGDIVDVMEKC.....DDGWFVGTSRRT.......RQFGTFPGNYVKPL
SH3P13  Novel      DQPCCRGLYDFEPENEGELGFKEGDIITLTNQI.....DENWYEGMLR.........GESGFFPINYVEVI
SH3P14  H74, m     TEVRVRALDYEGQEHDELSFKAGDELTKMEDED.....EQGWCKGRLDN........GQVGLYPANYVEAI
        H74, h     KGVRVRALYDYDGQEQDELSFKAGDELTKLGEED....EQGWCRGRLDS........GQLGLYPANYVEAI
SH3P15  Lyn        QGDIVVALYPYDGIHPDDLSFKKGEKMKVLEE......HGEWWKAKSLLT.......KKEGFIPSNYVAKL
SH3P16  Fyn        GVTLFVALYDYEARTEDDLSFHKGEKFQILNSS.....EGDWWEARSLTT.......GETGYIPSNYVAPV
SH3P17  Novel A    KVVYYRALYPFESRSHDEITIQPGDIVMVDESQTG...EPGWLGGELK.........GKTGWFPANYAEKI
              B    EGLQAQALYPWRAKKDNHLNFNKNDVITVLEQ......QDMWWFGEVQ.........GQKGWFPKSYVKLI
              C    GEEIAQVIASYTATGPEQLTLAPGQLILIRKKN.....PGGWWEGELQARGKK.....RQIGWFPANYVKLL
              D    AVCQVIAMYDYTAQNDDELAFNKGQIINVLNKE.....DPDWWKGEVN.........GQVGLFPSNYVKLT
```

FIG. 10A

```
SH3P18 Novel B  VGEEYIALYPYSSVEPGDLTFTEGEEILVTQK.....DGEWWTGSIG.....DRSGIFPSNYVKPK
           C    KPEIAQVTSAYVASGSEQLSLAPGQLILILKKN.....TSGWWQGELQARGKK.....RQKGWFPASYVKLL
           D    PVCQVIGMYDYAANNEDELSFSKGQLINVMNKD.....DPDWWQGEIN.....GVTGLFPSNYVLEE

      Src       GVTTFVALYDYESRTETDLSFKKGERLQIVNNT.....EGDWWLAHSLTT.....GQTGYIPSNYVAPS
```

FIG. 10B

CLONE 5	P3-6		QVKVFRALYTFEPRTPDELYFEEGDIIYITDM	DTNWWKGTS	GRTGLIPSNYVAEQ
CLONE 34	Crk LIKE		TGEEYIAVGDFTAQQVGDLTFKKGEILLVIEKK	PDGWWIAKDAK	GNEGLVPRTYLEPY
CLONE 40	Abl BINDING PROTEIN		YLEKVVAIYDYTKDKEDELSFQEGAIIYVIKKN	DDGWYEGVMN	GVTGLSPGNYVESI
CLONE 41	Nck LIKE	A	LNIPAFVKFAYVAIEREDELSLVKGSRVTVMEKC	SDGWWRGSYN	GQIGWFPSNYVLEE
	Nck LIKE	B	VLHVVQTLYPFSSVTEEELNEFEKGETMEVIEKPENDPEWWKCKNAR		GQVGLVPKNYVVVL
CLONE 45	Nck	A	EEVVVVAKFDYVAQQEQELDIKKNERLWLLDD	SKSWWRVRNSM	NKTGFVPSNYVERK
	Nck	B	LNMPAYVKFNYMAEREDELSLIKGTKVIVMEKIC	SDGWWRGSYN	GQVGWFPSNYVTEE
	Nck	C	VLHVVQALYPFSSSNDEELNFEKGDVMDVIEKPEN	DPEWWKCRKIN	GMVGLVPKNYVTVM
CLONE 53	NAB		DLFSYQALYSYIPQNDDELELRDGDIVDVMEKC	DDGWFVGTSRRT	KQFGTFPGNYVKPL
CLONE 55	NOVEL		QGRKERARYDLEAAQDNELTFKAGEIMTVLDDS	DPNWWKGETH	QGIGLFPSNFVTAD
CLONE 56	NOVEL		QGLCARALYDYQAADDTEISFDPENLITGIEVI	DEGWWRGYGPD	GHFGMFPANYVELI
CLONE 65	NOVEL	A	VLVNRALYPFEARNHDEMSFNSGDIIQVDEKTVG	EPGWLYGSFQ	GNFGWFPCNYVEKM
		B	VENLKAQALCSWTAKK DNHLNFSKHDIITVLEQQ	ENFWWFGEVH	GGRGWFPKSYVKII
		C	VGEEYIALYPYSSVEPGDLTFTEGEEILVTQK	DGEWWTGSIG	DRSGIFPSNYVKPK
		D	KPEIAQVTSAYVASGS EQLSLAPGQLILILKKN	TSGWWQGELQARGKKRQKGWFPASWVKLL	
		E	PVCQVIAMYDYAANNEDELSFSKGQLINVMNKD	DPDWWQGEIN	GVTGLFPSNYVKMT

FIG. 10C

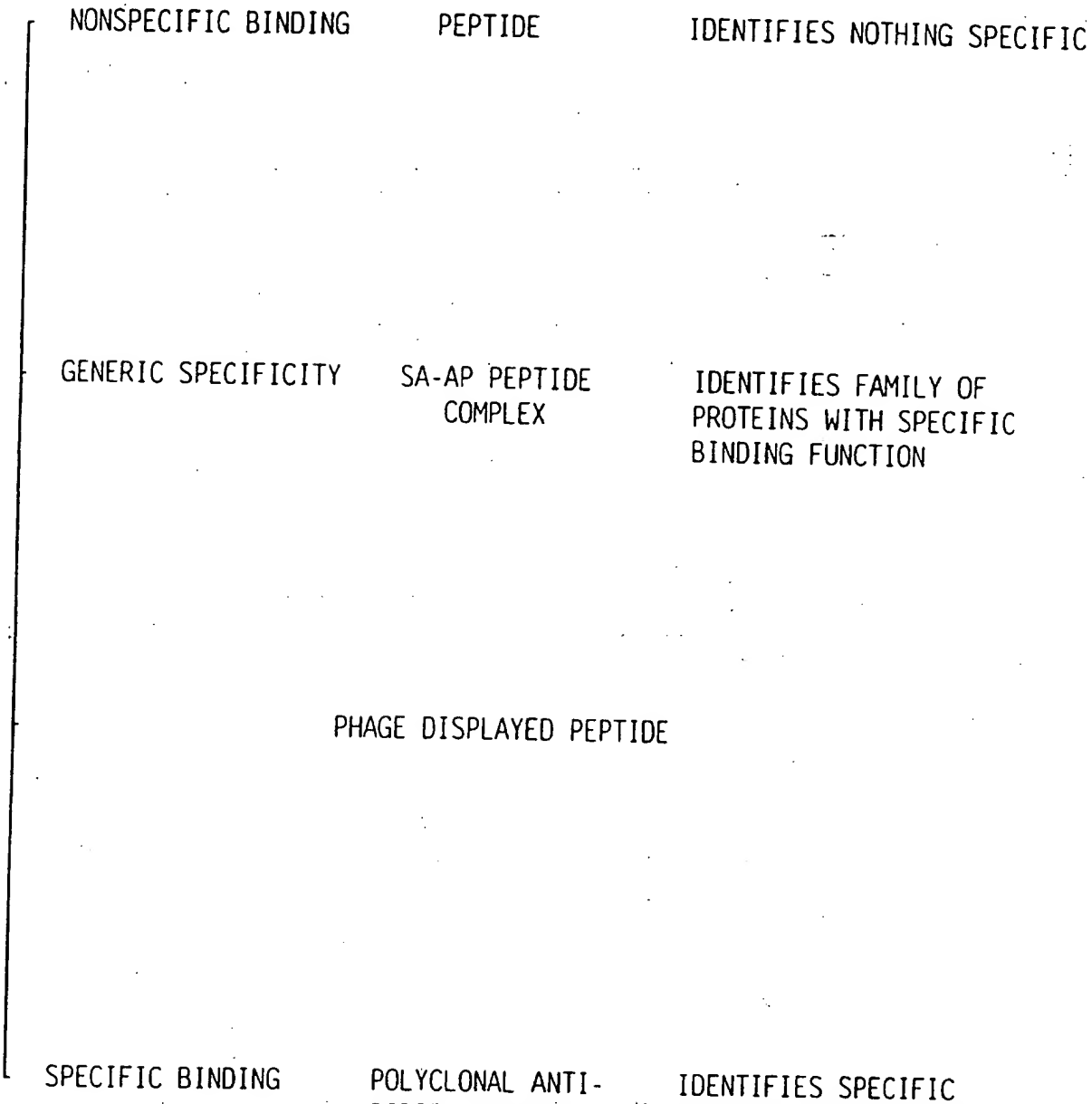
NONSPECIFIC BINDING
PEPTIDE
IDENTIFIES NOTHING SPECIFIC
GENERIC SPECIFICITY
SA-AP PEPTIDE COMPLEX
IDENTIFIES FAMILY OF PROTEINS WITH SPECIFIC BINDING FUNCTION
PHAGE DISPLAYED PEPTIDE
SPECIFIC BINDING
POLYCLONAL ANTI-BODIES HIGH AF- FINITY MABS SOME GST-FUSION PEPTIDES
IDENTIFIES SPECIFIC PROTEIN SCREENED FOR

FIG. 11

			SH3 DOMAIN CLONES														
PEPTIDE	NAME	SEQUENCE	5	11	12	13	14	18	34	40	41	45	46	53	55	56	65
SH3001	WBP-1	PGTPPPPYTVGPGY	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
TPPY	WBP-1	HGPTPPPPYTVGP	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
QPPY	WBP-2	YVQPPPPPYPGPM	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
YPPE	WBP-2	PGYPYPPPPEFY	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
WW005	WBP-1	PGTPAPPYTVGPGY	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
WW006	WBP-1	PGTPPAPYTVGPGY	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3002	K+ CHANNEL	DSGVRPLPPLPDPGV	–	–	–	+	+	–	++	–	–	–	–	–	–	–	–
bSH3003	K+ CHANNEL	VRPLPPLPEELPRPRRPPPED	–	–	+	+	+	–	++	–	–	–	–	–	–	–	–
bSH3004	M4 AChr	PPPALPPPPRPVADK	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3005	β1 ADRENERGIC	APAPPPGPPPAAAA	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3006	RasGap	GGGFPPLPPPPYLPPLG	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3007	MEK	SISPRPRPPGRPVSG	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3008	P Tyr PHOSP.	PPPEHIPPPPRPKRILE	–	–	+	–	–	++	–	+	–	–	–	–	–	–	–
bSH3009	Fak	KEGERALPSIPKLAN	–	–	–	+	+	–	–	–	–	–	–	–	–	–	–
bSH3010	c-Abl	SRLKPAPPPPPAASAG	–	–	–	–	–	+++	–	–	–	–	–	–	–	–	–
bSH3011	c-Cbl	QASLPPVPPRDLLLP	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3012	c-Cbl	PVPPTLRDLPPPPPPDRPYS	–	–	+++	+	++	–	+	–	–	–	–	–	–	–	–
bSH3013	$Ca2^+$ CHANNEL	SDQGRNLPGTPVPAS	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3014	$Ca2^+$ CHANNEL	RHSRRQLPPVPPKPRPLL	–	–	+	+	+	–	–	+	–	–	–	–	–	–	–
bSH3015	Nef	EKVGFPVTPQVPLRPMTY	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3016	MUS CADHERIN	PQPHRVLPTSPSDIA	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3017	AP2	ADFQPPYFPPPYQPTYPQS	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
bSH3018	ACTIN BINDING	SSAAPPPPPRRATPEK	–	+	+++	–	–	+++	–	+++	–	–	+	+++	–	–	–

FIG.12A

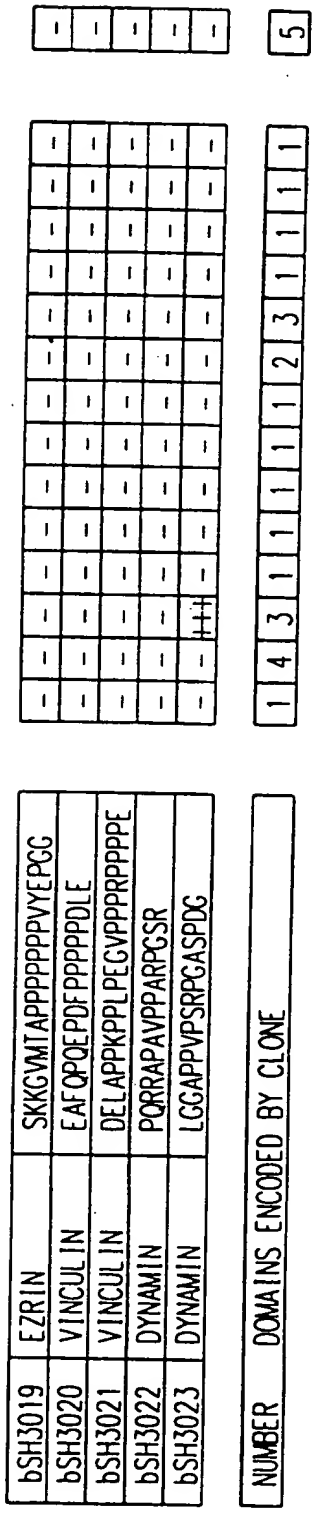
bSH3019 EZRIN SKKGVMTAPPPPPPPVYEPGG
bSH3020 VINCULIN EAFQPQEPDFPPPPPDLE
bSH3021 VINCULIN DELAPPKPPLPEGVPPPRPPPPE
bSH3022 DYNAMIN PQRRAPAVPPARPGSR
bSH3023 DYNAMIN LGGAPPVPSRPGASPDG
NUMBER DOMAINS ENCODED BY CLONE
1 4 3 1 1 1 1 1 2 3 1 1 1 1
5

FIG.12B

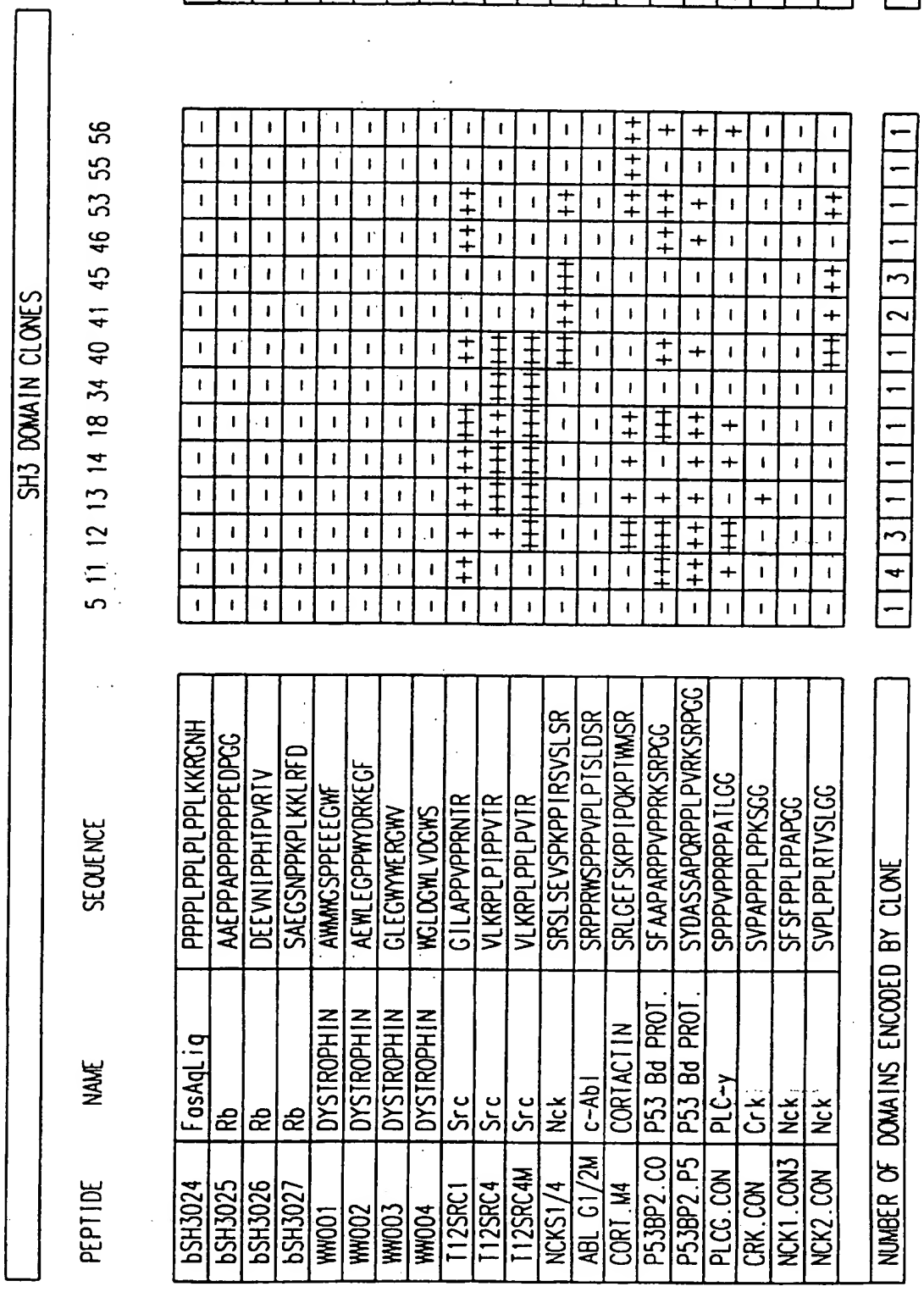
SH3 DOMAIN CLONES
PEPTIDE
NAME
SEQUENCE
5 11 12 13 14 18 34 40 41 45 46 53 55 56
65
bSH3024 FasAgLig PPPPLPPLPLPPLKKRGNH
bSH3025 Rb AAEPPAPPPPPPPEDPGG
bSH3026 Rb DEEVNIPPHIPVRTV
bSH3027 Rb SAEGSNPPKPLKKLRFD
WW001 DYSTROPHIN AWMMGSPPEEEGWF
WW002 DYSTROPHIN AEWLEGPPWYORKEGF
WW003 DYSTROPHIN GLEGWYWERGWV
WW004 DYSTROPHIN WGLDGWLVDGWS
T12SRC1 Src GILAPPVPPRNTR
T12SRC4 Src VLKRPLPIPPVTR
T12SRC4M Src VLKRPLPPLPVTR
NCKS1/4 Nck SRSLSEVSPKPPIRSVSLSR
ABL G1/2M c-Abl SRPPRWSPPPVPLPTSLDSR
CORT_M4 CORTACTIN SRLGEFSKPPIPQKPTWMSR
P53BP2.CO P53 Bd PROT. SFAAPARPPVPPRKSRPGG
P53BP2.P5 P53 Bd PROT. SYDASSAPQRPPLPVRKSRPGG
PLCG.CON PLC-γ SPPPVPPRPPATLGG
CRK.CON Crk SVPAPPPLPPKSGG
NCK1.CON3 Nck SFSFPPLPPAPGG
NCK2.CON Nck SVPLPPLRTVSLGG
NUMBER OF DOMAINS ENCODED BY CLONE
1 4 3 1 1 1 1 1 2 3 1 1 1 1
5

FIG.13

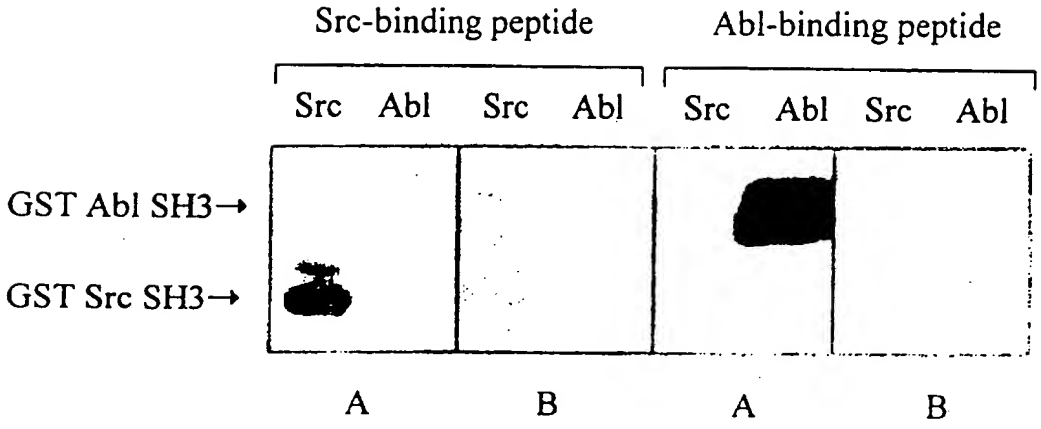
Src-binding peptide
Abl-binding peptide
Src Abl Src Abl
Src Abl Src Abl
GST Abl SH3→
GST Src SH3→
A
B
A
B

FIG.14

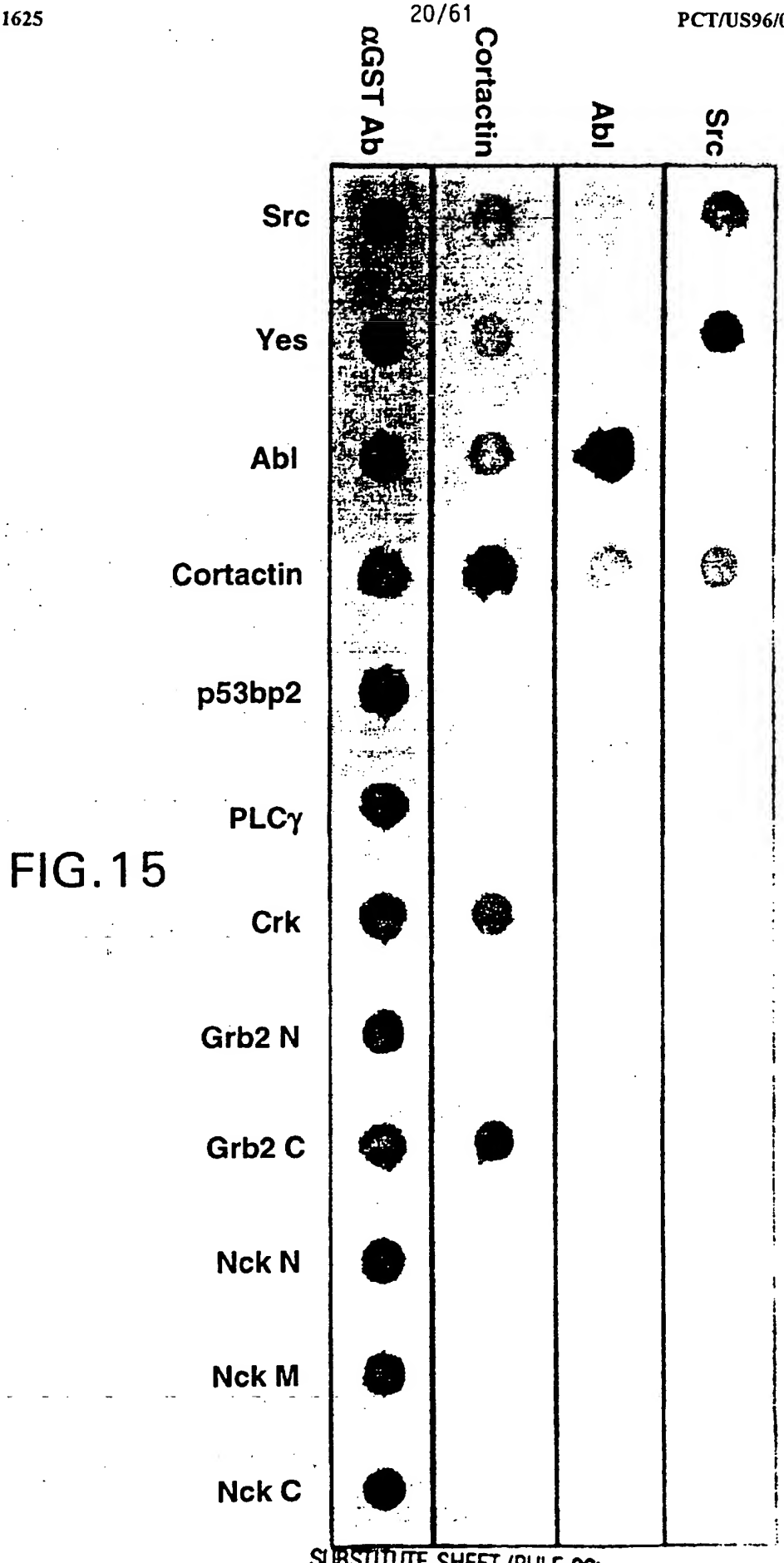
αGST Ab
Cortactin
Abl
Src
Src
Yes
Abl
Cortactin
p53bp2
PLCγ
Crk
Grb2 N
Grb2 C
Nck N
Nck M
Nck C

FIG.15

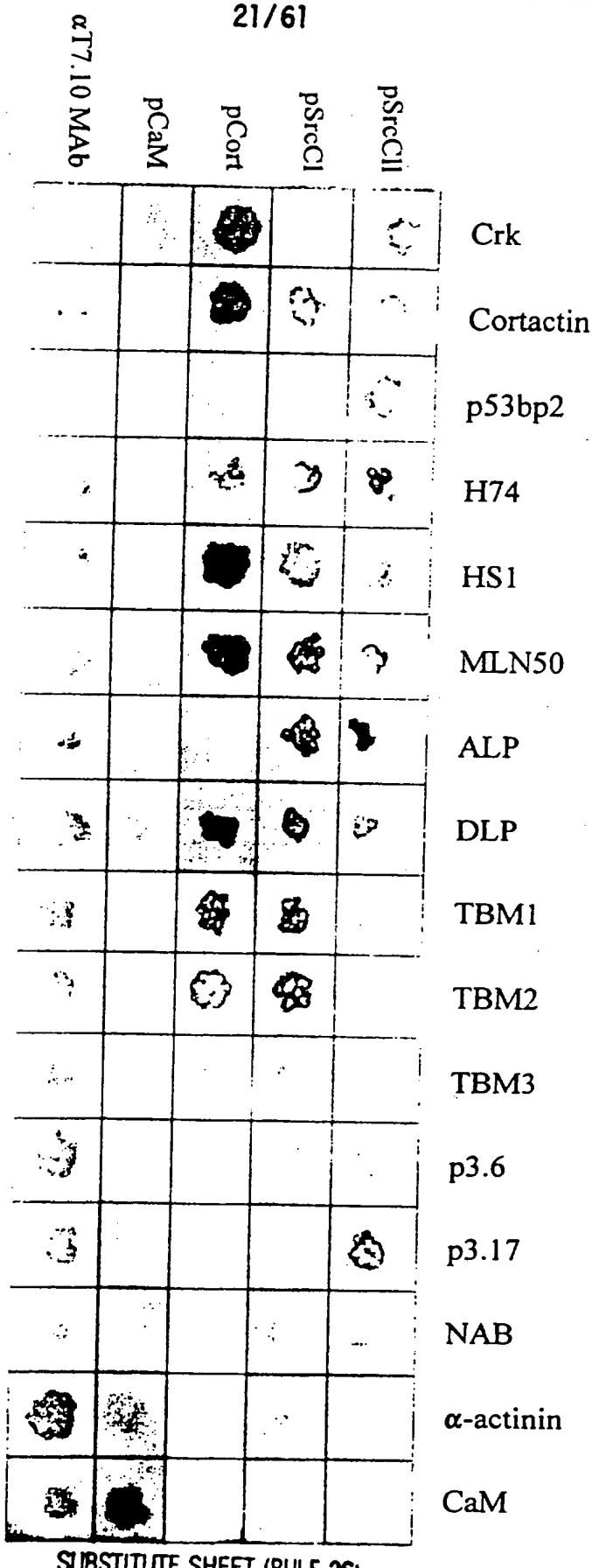
αT7.10 MAb
pCaM
pCort
pSrcCI
pSrcCII
Crk
Cortactin
p53bp2
H74
HS1
MLN50
ALP
DLP
TBM1
TBM2
TBM3
p3.6
p3.17
NAB
α-actinin
CaM

FIG.16

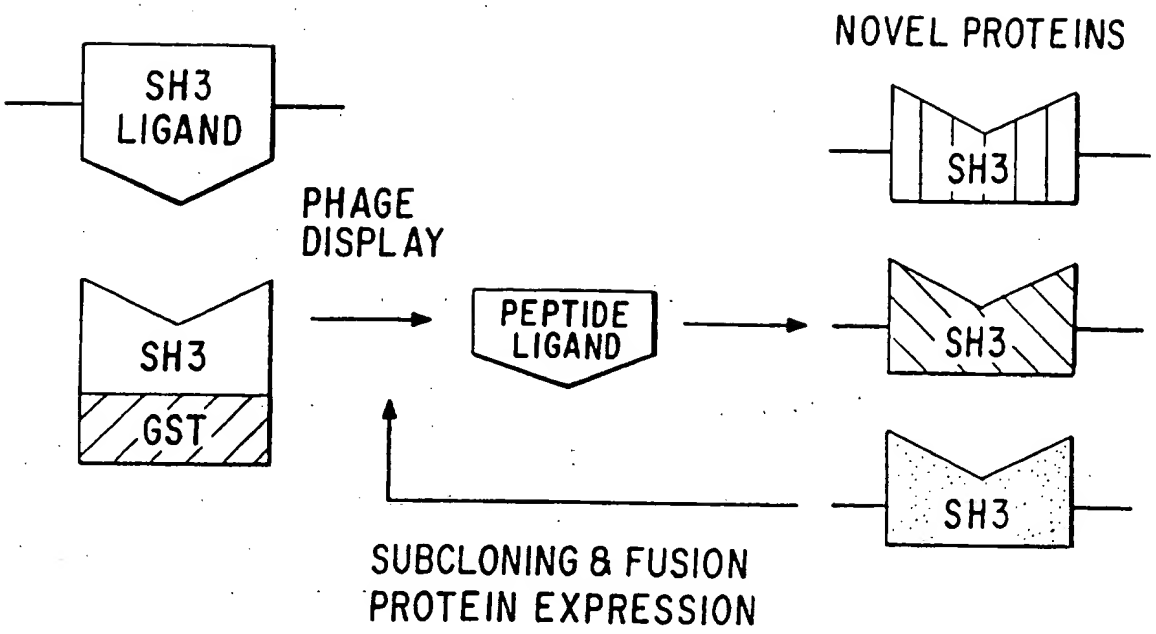
NOVEL PROTEINS
SH3 LIGAND
SH3
PHAGE DISPLAY
SH3
GST
PEPTIDE LIGAND
SH3
SH3
SUBCLONING & FUSION PROTEIN EXPRESSION

FIG. 17

```
1   GTGAATGCTG CAGACAGTGA CGGATGGACA CCACTGCATT
41  GTGCTGCCTC TTGCAACAGT GTCCACCTCT GCAAGCAGCT
81  GGTGGAAAGT GGAGCCGCTA TCTTTGCCTC CACCATCAGT
121 GACATTGAGA CTGCTGCAGA CAAGTGTGAA GAGATGGAAG
161 AGGGATACAT CCAGTGTTCC CAGTTTCTGT ATGGGGTACA
201 AGAGAAGCTG GGAGTGATGA ACAAAGGCAC CGTGTATGCT
241 TTGTGGGACT ACGAGGCCCA GAACAGCGAT GAGCTGTCCT
281 TCCATGAAGG GGATGCCATC ACCATCCTGA GGCGCAAAGA
321 TGAAAACGAG ACCGAGTGGT GGTGGGCTCG TCTTGGGGAC
361 CGGGAGGGCT ACGTGCCCAA AAACTTGCTG GGGTTGTATC
401 CACGGATCAA ACCCCGGCAG CGAACACTTG CCTGAACCCC
441 CTGGAGTACC ACAGTCTCGT TTGCTCCCAG GAGCTACTGG
481 AGGAGATCCC ACTGCCCTGG GAAAACTGAA GCTAGGATGG
521 TCTCCTGGTG CTCACTTTAG CAGACAGTGT CCACAATGTG
561 AATCCCACTT CCCAGGTGAG GCCCTCTCCA GGCTGCAGGA
601 GCTGG
```

FIG. 18

```
1   VNAADSDGWT PLHCAASCNS VHLCKQLVES GAAIFASTIS
41  DIETAADKCE EMEEGYIQCS QFLYGVQEKL GVMNKGTVYA
81  LWDYEAQNSD ELSFHEGDAI TILRRKDENE TEWWWARLGD
121 REGYVPKNLL GLYPRIKPRQ RTLA
```

FIG. 19

```
1   SGCARSGAAA ASAGLAPSCR VRVGLPRLSL VAPCSAMSKP
41  PPKPVKPGQV KVFRALYTFE PRTPDELYFE EGDIIYITDM
81  SDTSWWKGTC KGRTGLIPSN YVAEQAESID NPLHEAAKRG
121 NLSWLRECLD NRVGVNGLDK AGSTALYWAC HGGHKDIVEV
161 LFTQPNVELN QQNKLGDTAL HAAAWKGYAD IVQLLLAKGA
201 RTDLRNNEKK LALDMATNAA CASLLKKKQQ GTDGARTLSN
241 AEDYLDDEDS D
```

FIG 21

```
1     ..GAATTCAA GCTCGGGTTG CGCGCGGTCC GGAGCGGCCG
41    CGGCCAGCGC AGGCTTGGCG CCCAGTTGTC GTGTGCGTGT
81    GGGGCTCCCG CGGCTGAGCC TGGTCGCTCC GTGTAGCGCC
121   ATGTCCAAGC CACCTCCCAA ACCGGTCAAA CCAGGGCAAG
161   TTAAAGTCTT CAGAGCTCTA TATACATTTG AACCCAGAAC
201   TCCAGATGAA TTATACTTTG AAGAAGGAGA CATTATCTAC
241   ATCACTGACA TGAGTGATAC CAGCTGGTGG AAAGGGACAT
281   GCAAGGGCAG AACAGGACTG ATCCCGAGCA ACTATGTGGC
321   TGAGCAGGCA GAATCCATTG ACAATCCATT GCATGAAGCT
361   GCAAAAAGAG GCAACCTGAG CTGGTTGAGG GAGTGCTTGG
401   ACAACCGGGT GGGTGTGAAC GGCCTGGACA AAGCTGGAAG
441   CACAGCCCTG TACTGGGCCT GCCACGGTGG CCATAAAGAC
481   ATAGTGGAGG TTCTGTTTAC TCAGCCGAAT GTGGAGCTGA
521   ACCAGCAGAA TAAGCTGGGA GACACAGCTC TGCACGCGGN
561   TGCCTGGAAG GGTTATGCAG ACATTGTCCA GTTGCTACTG
601   GCAAAAGGTG CGAGGACAGA CTTGAGAAAC AATGAGAAGA
641   AGCTGGCCTT GGACATGGCC ACCAACGCTG CCTGTGCATC
681   GCTCCTGAAG AAGAAGCAGC AGGGAACAGA TGGGGCTCGA
721   ACGTTAAGCA ACGCCGAGGA CTACCTCGAT GACGAAGACT
761   CAGACTGATT CCCCCCGGGG CCGCTTTGAT TGTTGCCTAA
801   ACTTCTTTTG CTTTTGCCAT TCCGGAGCCT GGGTTGTTTG
841   CCAGAAGAGT ATTGATAACT GTTGCTTTTA AAGTCTGTAT
881   GAGCGCGACA CTGCTGCACT GTGATCTGTG AGGAGTCGTT
921   GTGAGGGTGG CTCATTCTCA CCCACGCCTT GNCAATAAGT
961   GAAGAGATAC TTTGTTGTAT AAAATACATA TATGCTCACC
1001  AGGGTAAAAT AAACGAAAAA AANTTATTTC TATTTATCAA
1041  GCTAAAAAAA AAAAGCTTGG GCCCTNTTCT ATAGTGTCAC
1081  CTAAATACTA GCTTGANCCG GNTGCTAACA AAGCCCGAAA
1121  GGAAGCTGAG TTGCTGCTGC CACCGNTGAG CAATAACTAG
1161  CATANCCCCT TGGGGCCTCT AAACGGGTCT TGAGGGGTTT
1201  TTNGNTGAAA GGAGGANCTA TTTCCGGATA ACCTGGNGTA
1241  ATAGGGAAGA GGCCCGNACC GATCGCCCTT CCCAACAGA
```

FIG. 20

```
1   .ACTCACGNC GGTGGAGTGG TACCGGATCG AATTCAAGCC GCATCACTGG
51  CACTGGACGC CAGGGCATCT TCCCTGCCAG CTACGTGCAG ATAAACCGAG
101 AGCCCCGGCT CAGGCTTTGT GATGATGGTC CCCAGCTCCC TGCATCACCT
151 AACCCGACAA CCACTGCTCA CCTAAGCAGC CACTCCCACC CCTCCTCAAT
201 ACCTGTGGAC CCCACTGACT GGGGAGGTCG AACCTCCCCT CGACGCTCCG
251 CCTTTCCCTT CCCCATCACC CTCCAGGAGC CCAGATCCCA AACCCAGAGT
301 CTCAATACCC CTGGACCAAC CCTGTCCCAT CCTCGAGCCA CCAGCCGTCC
351 CATAAACCTG GGACCCTCCT CCCCAAACAC AGAGATACAC TGGACTCCGT
401 ACCGGGCCAT GTACCAGTAC AGGCCCCAGA ATGAGGACGA GCTGGAACTT
451 CGAGAGGGGG ACCGTGTGGA TGTCATGCAG CAATGTGACG ATGGCTGGTT
501 TGTGGGTGTC TCCCGGCGAA CTCAGAAATT TGGGACATTC CCTGGAAATT
551 ATGTAGCCCC AGTGTGAGTG GTCTCCATGG CAGTTTGGAG CCAACGAGGA
601 TCGGGAGGGG AGCAGTAGCA CTATGGGAGG GAGAGAGGCC TTCCATAGCC
651 TCCTCCCCAG GACCTGTGCT CCCAGCTTCT GCAGAGACCC CAGCAACTTT
701 CCCTCCAAGC CTCCTTGAAG TCCGATTCCC ACCCCGCAAG TCACAGGCAT
751 TCCTTTGACA GCCCCCTTCA CCGCCCCTCA AATACAGACA TCTGCTTTCA
801 TGTGGGNAAA AAAAAAAAAT TAAAAGGTGG CCCTAT
```

FIG.22

```
1   RITGTGRQGI FPASYVQINR EPRLRLCDDG PQLPASPNPT
41  TTAHLSSHSH PSSIPVDPTD WGGRTSPRRS AFPFPITLQE
81  PRSQTQSLNT PGPTLSHPRA TSRPINLGPS SPNTEIHWTP
121 YRAMYQYRPQ NEDELELREG DRVDVMQQCD DGWFVGVSRR
161 TQKFGTFPGN YVAPV
```

FIG.23

```
1   MSVAGLKKQF HKATQKVSEK VGGAEGTKLD DDFKEMERKV
41  DVTSRAVMEI MTKTIEYLQP NPASRAKLSM INTMSKIRGQ
81  EKGPGYPQAE ALLAEAMLKF GRELGDDCNF GPALGEVGEA
121 MRELSEVKDS LDMEVKQNFI DPLQNLHDKD LREIQHHLKK
161 LEGRRLDFGY KKKRQGKIPD EELRQALEKF DESKEIAESS
201 MFNLLEMDIE QVSQLSALVQ AQLEYHKQAV QILQQVTVRL
241 EERIRQASSQ PRREYQPKPR MSLEFATGDS TQPNGGLSHT
281 GTPKPPGVQM DQPCCRALYD LEPENEGELA FKEGDIITLT
321 NQIDENWYEG MLHGQSGFFP INYVEILVAL PH
```

FIG.25

```
   1 TTNNNNYYMM SKYSKKGKKK KGKWMSGRTC GATTCAAGCC GACCAGCGGC
  51 GGCCCGGCGA CCCCAGCCGC CTCTCCGCAT CTGCATCTGC ATCTGCCGGC
 101 CGCGCAGCCT CCCGCATCCC ATCATGTCGG TGGCAGGGCT GAAGAAGCAG
 151 TTCCACAAAG CCACTCAGAA AGTGAGTGAG AAGGTGGGAG GAGCGGAAGG
 201 CACCAAGCTC GATGATGACT TCAAAGAGAT GGAGAGGAAA GTGGATGTCA
 251 CCAGCAGGGC TGTGATGGAG ATAATGACAA AAACGATTGA ATACCTCCAA
 301 CCCAATCCAG CTTCCAGGGC TAAGCTCAGT ATGATCAACA CCATGTCGAA
 351 AATCCGCGGC CAAGAGAAGG GGCCAGGCTA CCCTCAGGCG GAAGCACTGC
 401 TGGCAGAGGC CATGCTCAAG TTCGGCAGGG AGCTGGGTGA TGATTGCAAC
 451 TTTGGTCCTG CTCTCGGTGA GGTGGGAGAA GCCATGAGGG AGCTCTCGGA
 501 GGTCAAGGAC TCATTGGACA TGGAAGTGAA GCAGAATTTC ATCGACCCCC
 551 TTCAGAATCT TCATGACAAG GATCTGAGGG AGATTCAGCA TCATCTGAAA
 601 AAGCTGGAAG GCCGACGCTT AGACTTTGGT TATAAGAAGA AGCGACAAGG
 651 CAAGATTCCA GATGAAGAAC TCCGCCAAGC TCTGGAGAAA TTCGATGAGT
 701 CTAAAGAAAT CGCCGAGTCG AGCATGTTCA ACCTCTTGGA GATGGATATA
 751 GAACAGGTGA GCCAGCTCTC CGCACTTGTT CAGGCTCAGC TGGAGTACCA
 801 CAAGCAGGCA GTGCAGATCC TGCAGCAGGT CACTGTCAGA CTGGAAGAAA
 851 GAATAAGACA AGCTTCATCT CAGCCAAGAA GGGAATATCA GCCCAAACCA
 901 CGGATGAGCC TAGAGTTTGC CACTGGAGAC AGTACTCAGC CCAACGGGGG
 951 TCTCTCCCAC ACAGGCACAC CCAAACCTCC AGGTGTCCAA ATGGATCAGC
1001 CCTGCTGCCG AGCTCTGTAT GACTTGGAAC CTGAAAATGA AGGGGAATTG
1051 GCTTTTAAAG AGGGCGATAT CATCACACTC ACTAATCAGA TTGACGAGAA
1101 CTGGTATGAG GGGATGCTTC ATGGCCAGTC TGGCTTTTTC CCCATCAACT
1151 ATGTAGAAAT TCTGGTTGCT CTGCCCCATT AGGATCCTGT GCTGGCTGGC
1201 TCACCTCCTT CTGACCCAGA TAGTTAAGTT TAACCACTGC TTTGGTAATG
1251 CTGCTTCCAA TACATCACGA ATGCAGGCCG CAGTGGATGA GTCACCAAGC
1301 CCACACGTGC CCTGGGTTGA CCCGTGTGCT CCTCCAGGAG ACGCGGTGAT
1351 AGATGGTATC TTCCAAGGCC AGTGGGCCTG GTACATGCTT TAAAACACCA
1401 TCTGAGACTA GCCAGGAGTC CCAGAACTGG CTTCACAGTT CTCAGGAGGC
1451 TGTGGTTCCT GGTAACATGC CTGTGAACCA CATGGCAGAA AAACTCTCCT
1501 CACTGAAGAT ATTGTCTCTC ACCCAGGGGC CATCTCAAGG TCTCCAGTTC
1551 TCCATTTACA GAGGAGAAAG TCCTTTTTGT TGCACTTTCC CTTCCTAAAT
1601 ATGTGAGTCA CAGAATTGTT GGCAAAAACA TCCCCTCACC AGCAAGATGT
1651 CTGCTGGTTT AAGCAACTTG GTCTCTTGAT GCCATTAGCA AAAGTATTAA
1701 TTGTCCAAAG CACCTTTGTT CACTAATATC TATCTATCTA TCTATCTATC
1751 TATCTATCTA TCTATCTATC TATCTATCAT CTATCTACCT ACCTATCTAC
1801 CTATCATCTA TCTATCTATC ATCTATTATC TATCTATCTA TCTATCTATC
1851 NNTCNATCTA TCTATCTATC CATCTATCTA TCCATCATCT ATCTACCTAC
1901 CTATCTACTA TCCATCTATC TATCTATCCA TCATCTATCT ACCTACCTAT
1951 CTACTATCCA TCCATTTATC TATCTATCTA TCTATCTATC TATCTATCTA
2001 TCTCCCTCAT ACTTCTGAGA CATGGCCAGT TTTCTTCCCT CCCTGCTGTT
2051 AAGCACTTGG NAGATGAGGG GGGGGGTCCC ATTTNATTTC TGAGTGAGAT
2101 GGTGAGCAGG GTGTATGTTG GCTGTNNTNN GGGGGTGGCC CTA
```

FIG. 24

```
   1 CGGGCGCGGC GGGAGCCTGG TGGACCCTGC TTTGGCGGTA
  41 ATCATTGATC ATCGCAGATG CCCTCATATC CACTTTGGAT
  81 TCCTTGGATT CGGACAGACT CTGAACTGCT TTTCCCAGCA
 121 AAAGAGAAAG ATGTGGAAAG CCTCTGCAGG CCATGCTGTG
 161 TCCATCACGC AGGATGATGG AGGAGCTGAT GACTGGGAGA
 201 CTGATCCTGA TTTTGTGAAT GATGTGAGTG AAAAGGAGCA
 241 GAGATGGGGT GCTAAAACCG TGCAGGGATC GGGGCACCAG
 281 GAACACATCA ACATTCACAA GCTTCGAGAG AATGTCTTCC
 321 AAGAACACCA GACGCTCAAG GAGAAGGAGC TGGAAACGGG
 361 ACCCAAGGCT TCCCACGGCT ATGGCGGGAA GTTCGGTGTG
 401 GAGCAGGATA GGATGGACAG ATCAGCCGTG GGCCATGAGT
 441 ACCAGTCGAA GCTTTCCAAG CACTGCTCAC AAGTGGACTC
 481 GGTCCGGGGC TTCGGAGGCA AGTTCGGTGT CCAGATGGAC
 521 AGGGTGGATC AGTCTGCTGT AGGCTTTGAA TACCAGGGGA
 561 AGACTGAGAA GCATGCCTCC CAGAAAGACT ACTCTAGTGG
 601 CTTCGGTGGC AAATACGGTG TGCAAGCTGA CCGTGTAGAC
 641 AAGAGTGCCG TGGGCTTTGA CTACCAGGGC AAGACGGAGA
 681 AGCATGAGTC TCAGAAAGAT TACTCCAAAG GTTTTGGTGG
 721 CAAATATGGG ATTGACAAGG ACAAGGTGGA TAAAAGTGCT
 761 GTGGGCTTTG AGTATCAAGG CAAGACAGAG AAGCACGAAT
 801 CCCAGAAAGA CTATGTAAAA GGCTTTGGAG GAAAGTTTGG
 841 TGTGCAGACA GACAGACAGG ACAAGTGTGC CCTTGGCTGG
 881 GACCATCAGG AGAAGCTGCA GCTGCATGAA TCCCAAAAAG
 921 ACTATAAGAC TGGTTTCGGA GGCAAATTTG GTGTTCAGTC
 961 CGAGAGGCAG GACTCCTCCG CTGTGGGGTT TGATTACAAG
1001 GAGAGATTGG CCAAGCACGA GCCCCAGCAA GACTATGCCA
1041 AAGGATTCGG CGGGAAGTAT GGGGTGCAGA AGGATCGGAT
1081 GGACAAGAAT GCATCCACCT TTGAAGAAGT GGTCCAGGTG
1121 CCATCTGCCT ATCAGAAGAC TGTCCCCATT GAGGCCGTAA
1161 CCAGCAAAAC CAGTAATATC CGTGCTAACT TTGAAAACCT
1201 GGCAAAGGAG AGAGAGCAGG AGGACAGGCG GAAGGCAGAA
1241 GCCGAGAGAG CTCAGCGGAT GGCCAAAGAA AGACAGGAGC
1281 AGCAGGAGGC GCGCAGGAAG CTGGAAGAGC AAGCCAGAGC
1321 CAAGAAGCAG ACGCCCCCTG CATCCCCTAG TCCTCAACCA
1361 ATTGAAGACA GACCACCCTC CAGCCCCATC TATGAGGATG
1401 CAGCTCCGTT CAAGGCCGAG CCGAGCTACC GAGGTAGCGA
1441 ACCTGAGCCT GAGTACAGCA TCGAGGCCGC AGGCATTCCT
1481 GAGGCTGGCA GCCAGCAAGG CCTGACCTAT ACATCAGAGC
1521 CCGTGTACGA GACTACAGAG GCTCCTGGCC ACTATCAAGC
1561 AGAGGATGAC ACCTACGATG GGTATGAGAG TGACCTGGGC
1601 ATCACAGCCA TCGCCCTGTA TGACTACCAG GCTGCTGGCG
```

FIG. 26A

```
1641 ATGATGAGAT CTCCTTTGAC CCTGATGACA TCATCACCAA
1681 CATAGAAATG ATTGACGATG GCTGGTGGCG TGGGGTGTGC
1721 AAGGGCAGAT ACGGGCTCTT CCCAGCCAAC TATGTGGAGC
1761 TGCGGCAGTA GGGCTGCCAC CCAGAGCCTA CCGGCACCAG
1801 CACAGGGTTC ACACTACAGA GCATCTGCGT GTGTTTGAGT
1841 TGGTTTCTGC TTCCGTTTCT GTTTTTG
```

FIG. 26B

```
1   MWKASAGHAV SITQDDGGAD DWETDPDFVN DVSEKEQRWG
41  AKTVQGSGHQ EHINIHKLRE NVFQEHQTLK EKELETGPKA
81  SHGYGGKFGV EQDRMDRSAV GHEYQSKLSK HCSQVDSVRG
121 FGGKFGVQMD RVDQSAVGFE YQGKTEKHAS QKDYSSGFGG
161 KYGVQADRVD KSAVGFDYQG KTEKHESQKD YSKGFGGKYG
201 IDKDKVDKSA VGFEYQGKTE KHESQKDYVK GFGGKFGVQT
241 DRQDKCALGW DHQEKLQLHE SQKDYKTGFG GKFGVQSERQ
281 DSSAVGFDYK ERLAKHEPQQ DYAKGFGGKY GVQKDRMDKN
321 ASTFEEVVQV PSAYQKTVPI EAVTSKTSNI RANFENLAKE
361 REQEDRRKAE AERAQRMAKE RQEQEEARRK LEEQARAKKQ
401 TPPASPSPQP IEDRPPSSPI YEDAAPFKAE PSYRGSEPEP
441 EYSIEAAGIP EAGSQQGLTY TSEPVYETTE APGHYQAEDD
481 TYDGYESDLG ITAIALYDYQ AAGDDEISFD PDDIITNIEM
521 IDDGWWRGVC KGRYGLFPAN YVELRQ
```

FIG. 27

```
1    AAGCAGTCCT TCACCATGGT GGCCGACACT CCGGAAAACC TCCGCCTCAA
51   GCAACAGAGC GAGCTGCAGA GTCAGGTGCG CTACAAGGAG GAGTTTGAGA
101  AGAATAAGGG CAAAGGTTTC AGCGTGGTGG CAGACACGCC TGAGCTGCAG
151  AGAATCAAGA AGACCCAGGA CCAGATCAGC AATATCAAAT ACCATGAGGA
201  GTTTGAGAAG AGCCGCATGG GGCCCAGTGG AGGAGAAGGG GTGGAACCAG
251  AGCGCCGAGA AGCCCAGGAC AGCAGCAGCT ACCGGAGGCC CACAGAGCAG
301  CAGCAGCCGC AGCCTCACCA TATCCCGACC AGTGCCCCCG TGTACCAGCA
351  GCCCCAGCAG CAGCAGATGA CCTCGTCCTA TGGTGGGTAC AAGGAGCCAG
401  CAGCCCCTGT CTCCATACAG CGCAGTGCCC CAGGTGGCGG TGGGAAACGG
451  TACCGTGCAG TGTATGACTA CAGCGCTGCC GACGAGGACG AGGTCTCCTT
501  CCAGGATGGG GACACCATCG TCAATGTGCA GCAGATCGAT GACGGCTGGA
551  TGTACGGGAC CGTAGAGCGC ACCGGTGACA CGGGGATGCT GCCAGCCAAC
601  TACGTGGAGG CCATCTGAAC CCTGTGCCGC CCCGCCCTGT CTTCAATGCA
651  TTCCATGGCA TCACATCTGT CCTGGGGCCT GACCCGTCCA CCCTTCAGTG
701  TCTCTGTCTT TTAAGATCTT CAACTGCTTC TTTATCCCCG CCCCTCCAGC
751  TTATTTTACC ATCCCAAGCC TTGTTCTGCC CCTGTCATGG GCTCCTTCCT
801  CTGGCAGGTT TTCCCTTGGA CCAATCAACT GATTGATTTT TCTCTCTGGA
851  TGGAACAGGC TGGGCACTCT GGGGAGGGCA GGATTGTTCT TAGCTAGGTA
901  GACTCCCAGG GCTGGGCTGA ACTAGGAGAC CCACTAAGGA GATCAGTTTA
951  GACTGGGTGC AGTGGCAAAC ACCCTTAATT CCCAGCGAAG GGAGTCAGAG
1001 GCAGGCAGAT CTGTGACTTG GAAGCCAGCC TGGTCTACAT CGAGAGTTTC
1051 AGGACAGCCA GAGCTATGTA GTGAGGCCCT GTCTCGGAGG AAGAGTGGGG
1101 GTTGGTTAGC TCTCAGCTTC ACTTCCTGCC TTAGGCTCCT CAGAACCCCT
1151 GGCCCAGCTC CCCCAACTCC CTTCCTCCTA GAGGTGGGGT GAGCTGTGC
```

FIG. 28

```
1   KQSFTMVADT PENLRLKQQS ELQSQVRYKE EFEKNKGKGF SVVADTPELQ
51  RIKKTQDQIS NIKYHEEFEK SRMGPSGGEG VEPERREAQD SSSYRRPTEQ
101 QQPQPHHIPT SAPVYQQPQQ QQMTSSYGGY KEPAAPVSIQ RSAPGGGGKR
151 YRAVYDYSAA DEDEVSFQDG DTIVNVQQID DGWMYGTVER TGDTGMLPAN
201 YVEAI
```

FIG. 29

```
1    ATGGCGGTGA ACCTGAGCCG GAACGGGCCG GCGCTGCAGG AGGCCTACGT
51   GCGCGTAGTC ACCGAGAAAT CCCCGACCGA CTGGGCTCTT TTTACCTATG
101  AAGGCAACAG CAATGACATC CGTGTGGCTG GCACAGGAGA GGGAGGCCTG
151  GAGGAGCTGG TGGAAGAGCT CAACAGCGGG AAGGTGATGT ACGCCTTCTG
201  CAGGGTGAAG GACCCCAACT CCGGCCTGCC CAAGTTTGTC CTCATCAACT
251  GGACAGGAGA GGGTGTGAAT GATGTGCGGA AAGGAGCATG TGCCAACCAC
301  GTCAGCACCA TGGCCAACTT CCTGAAGGGT GCCCACGTGA CCATCAATGC
351  CCGGGCCGAG GAGGATGTGG AGCCTGAGTG CATCATGGAG AAGGTTGCCA
401  AGGCCTCTGG GGCCAACTAC AGCTTCCATA AGGAAAGCAC CTCCTTCCAG
451  GATGTAGGGC CGCAGGCCCC AGTGGGCTCT GTGTACCAGA AGACCAATGC
501  CATATCTGAG ATCAAGAGAG TCGGCAAGGA TAACTTCTGG GCCAAAGCTG
551  AGAAGGAAGA AGAGAACCGC CGCCTGGAGG AGAAGCGGCG TGCCGAAGAG
601  GAGCGGCAGC GGTTGGAGGA GGAGCGACGA GAGCGGGAGC TGCAGGAGGC
651  TGCCCGACGT GAGCAGCGCT ACCAGGAACA GCACAGATCA GCTGGAGCCC
701  CGAGCAGGAC AGGTGAGCCA GAGCAGGAAG CCGTTTCAAG GACCAGACAG
751  GAGTGGGAGT CTGCTGGGCA GCAGGCCCCA CACCCACGAG AGATTTTCAA
801  GCAGAAGGAA AGGGCAATGT CCACCACCTC TGTCACCAGC TCGCAGCCGG
851  GCAAGCTGAG GAGCCCCTTC CTGCAGAAGC AACTCACTCA ACCAGAAACC
901  TCCTACGGCC GAGAGCCCAC AGCTCCTGTC TCCCGGCCTG CAGCAGGTGT
951  CTGTGAGGAG CCAGCGCCTA GCACTCTGTC TTCTGCCCAG ACAGAAGAAG
1001 AACCTACATA TGAAGTACCC CCAGAGCAGG ACACCCTCTA TGAGGAACCA
1051 CCACTGGTAC AGCAGCAAGG GGCTGGCTCC GAACACATTG ACAACTACAT
1101 GCAGAGCCAG GGCTTCAGTG GACAAGGGCT GTGCGCCCGG GCCTTGTATG
1151 ACTACCAGGC AGCTGATGAC ACCGAGATCT CCTTTGACCC TGAGAACCTA
1201 ATCACAGGCA TCGAGGTGAT TGACGAAGGC TGGTGGCGAG GCTATGGGCC
1251 TGACGGCCAC TTTGGCATGT TTCCTGCCAA CTACGTGGAG CTCATAGAGT
1301 GA
```

FIG. 30

```
1   MAVNLSRNGP ALQEAYVRVV TEKSPTDWAL FTYEGNSNDI RVAGTGEGGL
51  EELVEELNSG KVMYAFCRVK DPNSGLPKFV LINWTGEGVN DVRKGACANH
101 VSTMANFLKG AHVTINARAE EDVEPECIME KVAKASGANY SFHKESTSFQ
151 DVGPQAPVGS VYQKTNAISE IKRVGKDNFW AKAEKEEENR RLEEKRRAEE
201 ERQRLEEERR ERELQEAARR EQRYQEQHRS AGAPSRTGEP EQEAVSRTRQ
251 EWESAGQQAP HPREIFKQKE RAMSTTSVTS SQPGKLRSPF LQKQLTQPET
301 SYGREPTAPV SRPAAGVCEE PAPSTLSSAQ TEEEPTYEVP PEQDTLYEEP
351 PLVQQQGAGS EHIDNYMQSQ GFSGQGLCAR ALYDYQAADD TEISFDPENL
401 ITGIEVIDEG WWRGYGPDGH FGMFPANYVE LIE
```

FIG. 31

```
1   MSVAGLKKQF YKASQLVSEK VGGAEGTKLD DDFKDMEKKV DVTSKAVAEV
51  LVRTIEYLQP NPASRAKLTM LNTVSKIRGQ VKNPGYPQSE GLLGECMVRH
101 GKELGGESNF GDALLDAGES MKRLAEVKDS LDIEVKQNFI DPLQNLCDKD
151 LKIEQHHLKK LEGRRLDFDY KKKRQGKIPD EELRQALEKF EESKEVAETS
201 MHNLLETDIE QVSQLSALVD AQLDYHRQAV QILEELADKL KRRVREASSR
251 PKREFKPRPR EPFELGELEQ PNGGFPCAPA PKITASSSFR SSDKPIRMPS
301 KSMPPLDQPS CKALYDFEPE NDGELGFREG DLITLTNQID ENWYEGMLHG
351 QSGFFPLSYV QVLVPLPQ
```

FIG.33

```
1   MAEMGSKGVT AGKIASNVQK KLTRAQEKVL QKLGKADETK DEQFEQCVQN
51  FNKQLTEGTR LQKDLRTYLA SVKAMHEASK KLSECLQEVY EPEWPGRDEA
101 NKIAENNDLL WMDYHQKLVD QALLTMDTYL GQFPDIKSRI AKRGRKLVDY
151 DSARHHYESL QTAKKKDEAK IAKAEEELIK AQKVFEEMNV DLQEELPSLW
201 NSRVGFYVNT FQSIAGLEEN FHKEMSKLNQ NLNDVLVSLE KQHGSNTFTV
251 KAQPSDNAPE KGNKSPSPPP DGSPAATPEI RVNHEPEPAS GASPGATIPK
301 SPSQPAEASE VVGGAQEPGE TAASEATSSS LPAVVVETFS ATVNGAVEGS
351 AGTGRLDLPP GFMFKVQAQH DYTATDTDEL QLKAGDVVLV IPFQNPEEQD
401 EGWLMGVKES DWNQHKELEK CRGVFPENFT ERVQ
```

FIG. 35

```
   1 TTNNCACTCA CCGTCCGTGG TNNNNSTMMC SGWYNKRNTK YRRKMSSKRW
  51 YKWKKCRRKS GCGGCGCCGA CCTGCGCGCG GAGGAAAGAA GTCGGTTCGG
 101 CGGCGCCGGC GGAAACCGGA GTTCGAGCGG GAGGCCTGAC GGCGGCAGGC
 151 GGCATGTCGG TGGCGGGGCT GAAGAAGCAG TTCTACAAGG CGAGCCAGCT
 201 GGTCAGCGAG AAGGTTGGTG GGGCCGAAGG GACCAAACTG GATGATGACT
 251 TTAAAGATAT GGAAAAGAAG GTGGATGTCA CCAGCAAGGC CGTGGCAGAG
 301 GTGCTGGTCA GAACCATAGA ATATCTGCAG CCTAACCCAG CCTCGAGAGC
 351 CAAGCTGACT ATGCTGAACA CCGTATCCAA GATCCGGGGC CAAGTGAAGA
 401 ACCCTGGCTA CCCACAGTCA GAGGGTCTGT TGGGAGAGTG CATGGTTCGC
 451 CATGGCAAGG AACTAGGTGG AGAGTCCAAC TTCGGTGATG CCCTGCTAGA
 501 TGCAGGTGAG TCCATGAAGC GCCTGGCTGA GGTGAAGGAC TCACTGGACA
 551 TCGAGGTCAA GCAGAACTTC ATTGACCCAC TACAGAACCT GTGTGACAAG
 601 GATCTGAAGG AGATCCAGCA CCACCTGAAG AAATTGGAGG GCCGCCGCCT
 651 TGACTTTGAC TACAAGAAGA AGCGCCAGGG CAAGATCCCC GATGAGGAGC
 701 TGCGCCAGGC CCTAGAGAAG TTCGAGGAGT CCAAGGAGGT GGCGGAGACC
 751 AGTATGCACA ACCTCCTGGA GACTGATATA GAGCAGGTGA GCCAGCTCTC
 801 GGCCCTGGTG GATGCCCAGC TGGACTACCA CCGGCAGGCA GTGCAGATCC
 851 TGGAGGAGCT GGCTGACAAG CTGAAGCGCA GGGTTCGGGA AGCCTCCTCA
 901 CGCCCCAAGC GGGAGTTCAA GCCCCGGCCC CGGGAGCCCT TTGAGCTTGG
 951 AGAGCTGGAG CAGCCCAATG GGGGATTCCC CTGTGCCCCA GCACCTAAGA
1001 TCACAGCCTC CTCATCATTT AGATCGTCAG ACAAGCCCAT CAGGATGCCC
1051 AGCAAGAGCA TGCCACCCCT GGACCAGCCA AGCTGCAAGG CGCTTTATGA
1101 TTTTGAGCCA GAGAATGATG GCGAGCTGGG CTTCCGTGAG GGGGACCTCA
1151 TCACGCTTAC CAACCAGATC GACGAGAACT GGTATGAGGG GATGCTGCAC
1201 GGCCAATCAG GCTTCTTCCC ACTCAGCTAC GTGCAGGTGC TGGTGCCTCT
1251 GCCTCAGTGA CTGGGCCTTT ACACCGCTGC CAGTCACAGT GCAGCAGCAG
1301 TCTAATGCCA AGGTGCTCTA GAAACACTAA TGTTCCTCCA GGGGGGACTC
1351 CTCCCCACTC CCTCAGCCCT GGGGCCCCCC TATCCTAAGA CTCGGAAAGG
1401 CCCACCCTGA GGTTCTATTG CCTTCCTGGT GGTATCAGCT TCCAGCTGTT
1451 TCAACCCTTC CCAGCCCGTT GCTGGCGATG GSCCNNYGCC CCCTCTCTAG
1501 GCTCTCTAGA GGCAGGCAGG TCCTTGGAAT CCCCAGCCTG CAAGCAGAGG
1551 CTGGCCAGCT CCCCAGCTCA GCACACAGAC ACACCTGGCA CCTGCTGCTC
1601 ATGAAGAAGT GCACAAGGCA CAAATGTGTA CACTTCCCAT GGGACCACAG
1651 ACCCAGCTCA GCTCTGTTGA AGACCAAGCA CAAAGGCCTT GAAGAGTGGA
1701 CATTCCCAGG TCCCTGGCAC CTTCCCTTGA GCCAGCTCCA TTGCTACTTA
1751 TTCATGTGAC TGAAGCTGAC CACAGGCAGC TGGCAGGTCC TTTTTTCAAC
1801 CAGCAGGCTA GGCTGGCCAT AGACCCAGCT CTGCCTCACC CTGCCATGTT
1851 CCAGTAATGG AGGCCTCCAG CCTGGGCTCT ATTACATTCT TCTCTACAGC
1901 TGCCCCATAA CCCGTGGCTT ATCCCTGGCA CGTGGGGCCA CACCCCACGC
1951 CCCCTGGATA GGCAACACTG TCCTGCTCCA GCCTGTGCTG ANATGAACTG
2001 TACTCCTAAT TTTTTTTTAA AAAAAAAGTA TTAAATNTCT CTTTCTATAT
2051 AAAANAAAGN TGGCCCTANN NGGA
```

FIG. 32

```
   1 CCTCACTCGC TCTCCCCGCG CACGCTCCGT CTCCGTCAGT CCCCTGAGCT
  51 GTTCTAGTGC GCGGCGTGGA GCCAGGGCTC AGGCTGGTGG AGCGGCCGGG
 101 GCTGGAGGCT GGGAGTGCGG CGCGCACGGC CTCCCCGCGC CATTATCCGC
 151 GCTCGCTTCG GGCGAGGCCG GCGCCAGGAT GGCAGAGATG GGGAGCAAGG
 201 GGGTGACGGC GGGGAAGATC GCCAGCAACG TACAGAAGAA GCTGACCCGA
 251 GCGCAGGAGA AGGTCCTGCA GAAACTGGGG AAGGCGGACG AGACGAAGGA
 301 CGAGCAGTTT GAGCAGTGTG TCCAGAACTT CAATAAGCAG CTGACAGAGG
 351 GTACCCGGCT GCAGAAGGAT CTTCGGACCT ATCTGGCTTC TGTTAAAGCG
 401 ATGCACGAAG CCTCCAAGAA GCTGAGTGAG TGTCTTCAGG AGGTGTACGA
 451 GCCCGAGTGG CCTGGCAGGG ATGAAGCAAA CAAGATTGCA GAGAACAATG
 501 ACCTACTCTG GATGGACTAC CACCAGAAGC TGGTGGACCA GGCTCTGCTG
 551 ACCATGGACA CCTACCTAGG CCAGTTCCCT GATATCAAGT CGCGCATTGC
 601 CAAGCGGGGG CGGAAGCTGG TGGACTATGA CAGTGCCCGG CACCACTATG
 651 AGTCTCTTCA AACCGCCAAA AAGAAGGATG AAGCCAAAAT TGCCAAGGCA
 701 GAAGAGGAGC TCATCAAAGC CCAGAAGGTG TTCGAGGAGA TGAACGTGGA
 751 TCTGCAGGAG GAGCTGCCAT CCCTGTGGAA CAGCCGTGTA GGTTTCTATG
 801 TCAACACGTT CCAGAGCATC GCGGGTCTGG AGGAAAACTT CCATAAAGAG
 851 ATGAGTAAGC TCAATCAGAA CCTCAATGAT GTCCTGGTCA GCCTAGAGAA
 901 GCAGCACGGG AGCAACACCT TCACAGTCAA GGCCCAACCC AGTGACAATG
 951 CCCCTGAGAA AGGGAACAAG AGCCCGTCAC CTCCTCCAGA TGGCTCCCCT
1001 GCTGCTACCC CTGAGATCAG AGTGAACCAT GAGCCAGAGC CGGCCAGTGG
1051 GGCCTCACCC GGGGCTACCA TCCCCAAGTC CCCATCTCAG CCAGCAGAGG
1101 CCTCCGAGGT GGTGGGTGGA GCCCAGGAGC CAGGGGAGAC AGCAGCCAGT
1151 GAAGCAACCT CCAGCTCTCT TCCGGCTGTG GTGGTGGAGA CCTTCTCCGC
1201 AACTGTGAAT GGGGCGGTGG AGGGCAGCGC TGGGACTGGA CGCTTGGACC
1251 TGCCCCCGGG ATTCATGTTC AAGGTTCAAG CCCAGCATGA TTACACGGCC
1301 ACTGACACTG ATGAGCTGCA ACTCAAAGCT GGCGATGTGG TGTTGGTGAT
1351 TCCTTTCCAG AACCCAGAGG AGCAGGATGA AGGCTGGCTC ATGGGTGTGA
1401 AGGAGAGCGA CTGGAATCAG CACAAGGAAC TGGAGAAATG CCGCGGCGTC
1451 TTCCCGGAGA ATTTTACAGA GCGGCTACAG TGACGGAGGA GCCTTCCGGA
1501 GTGTGAAGAA CCTTTCCCCC AAAGATGTGT G
```

FIG. 34

```
   1 GAATTCGTCG ACCCACGCGT CCGGTTTGAG CAGTGCGTCC
  41 AGAATTTCAA CAAGCAGCTG ACGGAGGGCA CCCGGCTGCA
  81 GAAGGATCTC CGGACCTACC TGGCCTCCGT CAAAGCCATG
 121 CACGAGGCTT CCAAGAAGCT GAATGAGTGT CTGCAGGAGG
 161 TGTATGAGCC CGATTGGCCC GGCAGGGATG AGGCAAACAA
 201 GATCGCAGAG AACAACGACC TGCTGTGGAT GGATTACCAC
 241 CAGAAGCTGG TGGACCAGGC GCTGCTGACC ATGGACACGT
 281 ACCTGGGCCA GTTCCCCGAC ATCAAGTCAC GCATTGCCAA
 321 GCGGGGGCGC AAGCTGGTGG ACTACGACAG TGCCCGGCAC
 361 CACTACGAGT CCCTTCAAAC TGCCAAAAAG AAGGATGAAG
 401 CCAAAATTGC CAAGGCCGAG GAGGAGCTCA TCAAAGCCCA
 441 GAAGGTGTTT GAGGAGATGA ATGTGGATCT GCAGGAGGAG
 481 CTGCCGTCCC TGTGGAACAG CCGCGTAGGT TTCTACGTCA
 521 ACACGTTCCA GAGCATCGCG GGCCTGGAGG AAAACTTCCA
 561 CAAGGAGATG AGCAAGCTCA ACCAGAACCT CAATGATGTG
 601 CTGGTCGGCC TGGAGAAGCA ACACGGGAGC AACACCTCCA
 641 CGGTCAAGGC CCAGCCCAGT GACAACGCGC CTGCAAAAGG
 681 GAACAAGAGC CCTTCGCCTC CAGATGGCTC CCCTGCCGCC
 721 ACCCCCGAGA TCAGAGTCAA CCACGAGCCA GAGCCGGCCG
 761 GCGGGGCCAC GCCCGGGGCC ACCCTCCCCA AGTCCCCATC
 801 TCAGCCAGCA GAGGCCTCGG AGGTGGCGGG TGGGACCCAA
 841 CCTGCGGCTG GAGCCCAGGA GCCAGGGGAG ACGGCGGCAA
 881 GTGAAGCAGC CTCCAGCTCT CTTCCTGCTG TCGTGGTGGA
 921 GACCTTCCCA GCAACTGTGA ATGGCACCGT GGAGGGCGGC
 961 AGTGGGGCCG GGCGCTTGGA CCTGCCCCCA GGTTTCATGT
1001 TCAAGGTACA GGCCCAGCAC GACTACACGG CCACTGACAC
1041 AGACGAGCTG CAGCTCAAGG CTGGTGATGT GGTGCTGGTG
1081 ATCCCCTTCC AGAACCCTGA AGAGCAGGAT GAAGGCTGGC
1121 TCATGGGCGT GAAGGAGAGC GACTGGAACC AGCACAAGGA
1161 GCTGGAGAAG TGCCGTGGCG TCTTCCCCGA GAACTTCACT
1201 GAGAGGGTCC CATGACGGCG GGGCCCAGGC AGCCTCCGGG
1241 CGTGTGAAGA ACACCTCCTC CCGAAAAATG TGTGGTTCTT
1281 TTTTTTGTTT TGTTTTCGTT TTTCATCTTT TGAAGAGCAA
1321 AGGGAAATCA AGAGGAGACC CCCAGGCAGA GGGGCGTTCT
1361 CCCAAAGATT AGGTCGTTTT CCAAAGAGCC GCGTCCCGGC
1401 AAGTCCGGCG GAATTCACCA GTGTCCTGAA GCTGCTGTGT
1441 CCTCTAGTTG AGTTCTGGCG CCCCTGCCTG TGCCCGCATG
1481 TGTGCCTGGC CGCAGGGCGG GGCTGGGGGC TGCCGAGCCA
1521 CCATGCTTGC CTGAAGCTTC GGCCGCGCCA CCCGGGCAAG
1561 GGTCCTCTTT TCCTGGCAGC TGCTGTGGGT GGGGCCCAGA
1601 CACCAGCCTA ACCTGGCTCT GCCCCGCAGA CGGTCTGTGT
1641 GCTGTTTGAA AATAAATCTT AGTGTTCAAA ACAAAATGAA
1681 ACAAAAAAAA TGATAAAAAA AAAAAAAAAA AAAAAAAAAA
1721 AAAAGGGCGG CCGC
```

FIG. 36

```
1   EFVDPRVRFE QCVQNFNKQL TEGTRLQKDL RTYLASVKAM
41  HEASKKLNEC LQEVYEPDWP GRDEANKIAE NNDLLWMDYH
81  QKLVDQALLT MDTYLGQFPD IKSRIAKRGR KLVDYDSARH
121 HYESLQTAKK KDEAKIAKAE EELIKAQKVF EEMNVDLQEE
161 LPSLWNSRVG FYVNTFQSIA GLEENFHKEM SKLNQNLNDV
201 LVGLEKQHGS NTSTVKAQPS DNAPAKGNKS PSPPDGSPAA
241 TPEIRVNHEP EPAGGATPGA TLPKSPSQPA EASEVAGGTQ
281 PAAGAQEPGE TAASEAASSS LPAVVVETFP ATVNGTVEGG
321 SGAGRLDLPP GFMFKVQAQH DYTATDTDEL QLKAGDVVLV
361 IPFQNPEEQD EGWLMGVKES DWNQHKELEK CRGVFPENFT
401 ERVP
```

FIG. 37

```
1   MWKSVVGHDV SVSVETQGDD WDTDPDFVND ISEKEQRWGA KTIEGSGRTE
51  HINIHQLRNK VSEEHDILKK KELESGPKAS HGYGGQFGVE RDRMDKSAVG
101 HEYVADVEKH SSQTDAARGF GGKYGVERDR ADKSAVGFDY KGEVEKHASQ
151 KDYSHGFGGR YGVEKDKRDK AALGYDYKGE TEKHESQRDY AKGFGGQYGI
201 QKDRVDKSAV GFNEMEAPTT AYKKTTPIEA ASSGARGLKA KFESLAEEKR
251 KREEEEKAQQ MARQQQERKA VVKMSREVQQ PSMPVEEPAA PAQLPKKISS
301 EVWPPAESHL PPESQPVRSR REYPVPSLPT RQSPLGNHLE DNEEPPALPP
351 RTPEGLQVVE EPVYEAAPEL EPEPEPDYEP EPETEPDYED VGELDRQDED
401 AEGDYEDVLE PEDTPSLSYQ AGPSAGAGGA GISAIALYDY QGEGSDELSF
451 DPDDIITDIE MVDEGWWRGQ CRGHFGLFPA NYVKLL
```

FIG. 39

```
1   MAGNFDSEER SSWYWGRLSR QEAVALLQGQ RHGVFLVRDS STSPGDYVLS
51  VSENSRVSHY IINSSGPRPP VPPSPAQPPP GVSPSRLRIG DQEFDSLPAL
101 LEFYKIHYLD TTTLIEPVAR SRQGSGVILR QEEAEYVRAL FDFNGNDEED
151 LPFKKGDILR IRDKPEEQWW NAEDSEGKRG MIPVPYVEKY RPASASVSAL
201 IGGNQEGSHP QPLGGPEPGP YAQPSVNTPL PNLQNGPIYA RVIQKRVPNA
251 YDKTALALEV GELVKVTKIN VSGQWEGECN GKRGHFPFTH VRLLDQQNPD
301 EDFS
```

FIG. 41

```
1    GGATCCCCGG AGCCGGTCCG CTGGGCGGGG CGCAGGGCTG GAGGGGCGCG
51   CGTGCCGGCG GCGGCCCAGC GTGAAAGCGC GGAGGCGGCC ATGGCGGGCA
101  ACTTCGACTC GGAGGAGCGG AGTAGCTGGT ACTGGGGCCG CCTGAGCCGG
151  CAGGAGGCGG TGGCGCTATT GCAGGGCCAG CGGCACGGGG TGTTCCTGGT
201  GCGGGACTCG AGCACCAGCC CCGGGGACTA TGTGCTTAGC GTCTCCGAAA
251  ACTCGCGCGT CTCCCACTAC ATCATCAACA GCAGCGGCCC GCGCCCTCCA
301  GTGCCTCCGT CGCCCGCTCA GCCTCCGCCG GGAGTGAGTC CCTCCAGGCT
351  CCGAATAGGA GATCAAGAAT TTGATTCATT GCCTGCTTTA CTGGAATTCT
401  ACAAAATACA CTATTTGGAC ACTACAACAT TGATAGAACC AGTGGCCAGA
451  TCAAGGCAGG GTAGTGGAGT GATTCTCAGG CAGGAGGAGG CAGAGTATGT
501  GCGGGCCCTG TTTGACTTTA ATGGGAATGA TGAAGAAGAT CTTCCCTTTA
551  AGAAAGGAGA CATCCTGAGA ATCCGGGATA AGCCTGAAGA CGAGTGGTGG
601  AATGCAGAGG ACAGCGAAGG AAAGAGGGGG ATGATTCCTG TCCCTTACGT
651  GGAGAAGTAT AGACCTGCCT CCGCCTCAGT ATCGGCTCTG ATTGGAGGTA
701  ACCAGGAGGG TTCCCACCCA CAGCCACTGG GTGGGCCGGA GCCTGGGCCC
751  TATGCCCAAC CCAGCGTCAA CACTCCGCTC CCTAACCTCC AGAATGGGCC
801  CATTTATGCC AGGGTTATCC AGAAGCGAGT CCCTAATGCC TACGACAAGA
851  CAGCCTTGGC TTTGGAGGTC GGTGAGCTGG TAAAGGTTAC GAAGATTAAT
901  GTGAGTGGTC AGTGGGAAGG GGAGTGTAAT GGCAAACGAG GTCACTTCCC
951  ATTCACACAT GTCCGTCTGC TGGATCAACA GAATCCCGAT GAGGACTTCA
1001 GCTGAGTATA GCTCGACAGT TTGCTGACAG ATGGAACAAT CTGTTTTCCC
1051 CCAATTGCCA TCTATACAAT TTTCTTACAG GTGTCAAAGC AGTCTAGTTT
1101 ATATAAGCAT TCTGTTACCT GGGATCTTTT TTAAGACTGA ACTACTCCAT
1151 TCTCACTTGT ATTTACCATA TTCAGGGTAC GAAACCGGAG GGCTTATGTG
1201 GTTAACTTCT GAGTTGGCAG TTTTAGGTGG TAGTGGCCGT GCCTGTATGA
1251 GAAGACGTAA ATACATTGCC TCCTTTCTTT TGGGCAAAAC AGATCA
```

FIG. 40

```
1   MSSECDVGSS KAVVNGLASG NHGPDKDMDP TKICTGKGTV TLRASSSYRG
51  TPSSSPVSPQ ESPKHESKSD EWKLSSSADT NGNAQPSPLA AKGYRSVHPS
101 LSADKPQGSP LLNEVSSSHI ETDSQDFPPT SRPSSAYPST TIVNPTIVLL
151 QHNREQQKRL SSLSDPASER RAGEQDPVPT PAELTSPGRA SERRAKDASR
201 RVVRSAQDLS DVSTDEVGIP LRNTERSKDW YKTMFKQIHK LNRDDDSDVH
251 SPRYSFSDDT KSPLSVPRSK SEMNYIEGEK VVKRSATLPL PARSSSLKSS
301 PERNDWEPLD KKVDTRKYRA EPKSIYEYQP GKSSVLTNEK MSRDISPEEI
351 DLKNEPWYKF FSELEFGRPS SAVSPTPDIT SEPPGYIYSS NFHAVKRESD
401 GTPGGLASLE NERQIYKSVL EGGDIPLQGL SGLKRPSSSA STKDSESPRH
451 FIPADYLEST EEFIRRRHDD KEKLLADQRR LKREQEEADI AARRHTGVIP
501 THHQFITNER FGDLLNIDDT AKRKSGLEMR PARAKFDFKA QTLKELPLQK
551 GDVVYIYRQI DQNWYEGEHH GRVGIFPRTY IELLPPAEKA QPRKLAPVQV
601 LEYGEAIAKF NFNGDTQVEM SFRKGERITL LRQVDENWYE GRIPGTSRQG
651 IFPITYVDVL KRPLVKTPVD YIDLPYSSSP SRSATVSPQA SHHSLSAGPD
701 LTESEKNYVQ PQAQQRRVTP DRSQPSLDLC SYQALYSYVP QNDDELELRD
751 GDIVDVMEKC DDGWFVGTSR RTRQFGTFPG NYVKPLYL
```

FIG. 43

```
   1 CAGCCGCTGG AGGGGGCGCC TGGTGTAGAT GTGAAAAGCC GTAACCAGGA
  51 ACCAGTAAAG ATGTGGAAGT CTGTAGTGGG GCATGATGTA TCGGTTTCCG
 101 TGGAGACCCA GGGTGATGAC TGGGATACAG ACCCTGACTT TGTGAATGAC
 151 ATCTCCGAGA AGGAGCAACG GTGGGGAGCC AAGACCATTG AGGGCTCTGG
 201 ACGCACAGAG CACATCAACA TCCACCAGCT GAGGAACAAA GTGTCAGAGG
 251 AGCACGACAT CCTCAAGAAG AAGGAGCTGG AATCGGGGCC TAAGGCATCC
 301 CATGGCTATG GCGGTCAGTT TGGAGTGGAG AGAGACCGGA TGGACAAGAG
 351 TGCCGTGGGC CACGAGTATG TTGCTGATGT GGAGAAACAC TCATCTCAGA
 401 CTGATGCSGC CAGAGGCTTT GGGGGCAAAT ATGGAGTTGA GAGGGACCGG
 451 GCAGACAAGT CAGCGGTGGG CTTTGACTAC AAAGGAGAAG TGGAAAAGCA
 501 TGCATCTCAG AAAGATTACT CTCATGGCTT TGGTGGCCGC TACGGGGTAG
 551 AGAAGGATAA ACGGGACAAA GCAGCCCTGG GATACGACTA CAAAGGAGAG
 601 ACGGAGAAGC ACGAGTCTCA GAGAGATTAT GCCAAGGGCT TTGGTGGCCA
 651 ATATGGAATC CAGAAAGACC GAGTGGATAA GAGTGCTGTT GGCTTCAATG
 701 AAATGGAGGC CCCAACCACG GCGTATAAGA AGACAACACC CATAGAAGCT
 751 GCTTCCAGTG GTGCCCGTGG GCTGAAGGCA AAATTTGAGT CCCTGGCTGA
 801 GGAGAAGAGG AAGCGAGAGG AAGAAGAGAA GGCACAGCAG ATGGCCAGGC
 851 AGCAACAGGA GCGAAAGGCT GTGGTAAAGA TGAGCCGAGA AGTCCAGCAG
 901 CCATCCATGC CTGTGGAAGA GCCAGCGGCA CCAGCCCAGT TGCCCAAGAA
 951 GATCTCCTCA GAGGTCTGGC CTCCAGCAGA GAGTCACCTA CCGCCAGAGT
1001 CTCAGCCAGT GAGAAGCAGA AGGGAATACC CTGTGCCCTC TCTGCCCACG
1051 AGGCAGTCTC CATTGCAGAA TCACTTGGAG GACAACGAGG AGCCCCCAGC
1101 TCTGCCCCCT AGGACCCCAG AAGGCCTCCA GGTGGTGGAA GAGCCAGTGT
1151 ACGAAGCAGC ACCCGAGCTG GAGCCGGAGC CAGAGCCTGA CTATGAGCCA
1201 GAGCCAGAGA CAGAGCCTGA CTATGAGGAT GTTGGGGAGT TAGATCGGCA
1251 GGATGAGGAT GCAGAGGGAG ACTATGAGGA TGTGCTGGAG CCCGANGACA
1301 CCCCTTCTCT GTCCTACCAA GCTGGACCCT CAGCTGGGGC TGGTGGTGCG
1351 GGGATCTCTG CTATAGCCCT GTATGATTAC CAAGGAGAGG GAAGCGATGA
1401 GCTTTCCTTT GATCCAGATG ACATCATCAC TGACATTGAG ATGGTGGATG
1451 AAGGCTGGTG GCGGGGCCAA TGCCGTGGCC ACTTTGGACT TTTCCCTGCA
1501 AACTATGTCA AGCTCCTCTA ATGACCAGCC CATTGTCTTC CGACTTCCCG
1551 AATTCGAAGC TGCTCTGCCT CCCTCTTCCC ACTCCATGGT ACTGCTGCAA
1601 GGACCTGGCT GAACATCATG AGATGCCTGA AGTTCTGGCA GTCTGTCTCC
1651 CGCCTCTTTA AGAGCTTTAG GTAGAATCGC TCCAGGTGGG GGTGGGGGTG
1701 GGGGTGGGAT CCCTCTGTCC CTCTGTGACC ACTCTTCCCT GAGGTAGCTC
1751 ATGAAATCAT CTTGCAGACC TGCCTCCTTC AGCCGCACCC CAGCTCTGCC
1801 AACCTTGCTC TAGAGTGCTG GGATTCCCTT GCCCCGACCC TGGGTGCCAG
1851 CCTAGAGGGG AGGCTCTCAC AGGGCTGCCT GATTCGCCCT GTTGTGCTTT
1901 TGCTCATTTT TCTTCCCTTA GCAGACAAAT TGGAACTGCC CTTCTGTTTA
1950 GTCCTAAAAC TGAAAATAAA ATGAGACTGT GGCTAAAAAA AAAAAAAAAA
2003 AAA
```

FIG. 38

```
   1 CCTCACCGNN CCTGGTGTAG GTACCGGATC GAATTCAAGC GAAAAACAGA
  51 GCGGGGCTGA CTGTAGCGTG GAGCGCGAGC CGGGCTGGAC GCGCGCAAGC
 101 CCTTGCCGGG GACCCGCGAG GCAAGCAGTC TCCCTGTGGA GCGTCGTCCT
 151 CCATCCCTGT AAGCACCGTT ACAGAGAATG AAACAAGGGC AGAAGTTACA
 201 GAGCCCGTGA GGCATCTTCA AATAGAAGAC TGGAGACTAG AAASAGAATA
 251 TTGCCAGGAG TTGGCATCCA TTGGAAGACC TTGAGATCCT CTCAGCTCAC
 301 AACTCCAGGA CCGATGCATC TTCCCACCAC CTTGAAGCAC TGAGCCCTCC
 351 AGAGCTGCAT CTGGGAAGAC TCGCCTGCCT CCAGCATGAG TTCTGAATGT
 401 GATGTTGGAA GCTCTAAAGC TGTGGTGAAT GGCTTGGCAT CTGGCAACCA
 451 TGGACCAGAC AAAGACATGG ACCCTACCAA AATCTGCACT GGGAAAGGAA
 501 CAGTGACTCT TCGGGCCTCG TCTTCCTACA GGGGAACCCC AAGCAGCAGC
 551 CCTGTGAGCC CCCAGGAATC TCCGAAGCAT GAAAGCAAGT CAGATGAATG
 601 GAAACTTTCT TCCAGTGCAG ATACCAATGG CAACGCCCAG CCCTCCCCAC
 651 TTGCTGCCAA GGGCTATAGA AGTGTGCATC CCAGCCTTTC TGCTGACAAG
 701 CCCCAGGGCA GTCCTTTACT AAACGAAGTT TCTTCTTCCC ACATTGAAAC
 751 CGATTCCCAA GACTTCCCTC CAACAAGCAG ACCTTCGTCT GCCTACCCCT
 801 CCACCACCAT CGTCAACCCT ACCATTGTGC TCCTGCAGCA CAATCGAGAG
 851 CAGCAAAAGC GACTCAGTAG TCTTTCAGAT CCTGCCTCAG AGAGAAGAGC
 901 GGGTGAGCAG GACCCAGTAC CAACCCCAGC AGAACTCACT TCGCCCGGCA
 951 GGGCTTCTGA GAGAAGGGCA AAGGATGCTA GCAGACGGGT GGTGAGGAGC
1001 GCACAGGACC TGAGCGATGT GTCTACAGAT GAAGTGGGCA TTCCACTCCG
1051 GAATACCGAG CGATCGAAAG ACTGGTACAA AACTATGTTT AAACAGATCC
1101 ACAAACTGAA CAGAGATGAT GATTCTGATG TCCATTCCCC TCGATACTCC
1151 TTCTCTGATG ACACAAAGTC TCCCCTTTCT GTGCCTCGCT CAAAAAGTGA
1201 GATGAACTAC ATCGAAGGGG AGAAAGTGGT TAAGAGGTCC GCCACACTCC
1251 CCCTCCCAGC CCGCTCTTCC TCACTCAAGT CCAGCCCGGA AAGAAACGAC
1301 TGGGAGCCCC TAGATAAGAA AGTGGATACG AGAAAATACC GAGCAGAGCC
1351 CAAAAGCATT TACGAATATC AGCCGGGCAA GTCTTCGGTC CTGACCAATG
1401 AGAAGATGAG TCGGGATATA AGCCCAGAAG AGATAGATTT AAAGAATGAA
1451 CCTTGGTATA AATTCTTTTC GGAATTGGAG TTTGGGAGAC CGAGCTCAGC
1501 AGTCAGCCCG ACTCCAGACA TTACGTCAGA GCCTCCTGGA TATATCTATT
1551 CTTCCAACTT CCATGCAGTG AAGAGAGAAT CGGACGGGAC CCCCGGGGGT
1601 CTCGCTAGCT TGGAGAATGA GAGGCAGATC TATAAGAGTG TCTTGGAAGC
1651 TGGCGACATC CCTCTTCAGG GCCTCAGTGG GCTCAAGCGA CCTTCCAGCT
1701 CAGCTTCCAC TAAAGATTCA GAGTCACCAA GACATTTTAT ACCAGCTGAT
1751 TACTTGGAGT CCACAGAAGA ATTTATTCGG AGACGGCACG ATGATAAAGA
1801 GAAACTTTTA GCGGACCAGA GACGACTTAA GCGCGAGCAA GAAGAGGCCG
1851 ATATTGCAGC TCGCCGCCAC ACAGGTGTCA TCCCGACTCA TCATCAGTTT
1901 ATCACTAATG AGCGCTTTGG GGACCTCCTC AATATAGATG ATACGGCCAA
1951 AAGGAAATCT GGGTTAGAGA TGAGACCTGC TCGAGCCAAA TTTGACTTTA
2001 AAGCCCAGAC CCTGAAGGAG CTGCCTCTGC AGAAGGGAGA CGTTGTTTAC
2051 ATCTACAGAC AGATTGACCA GAACTGGTAT GAAGGTGAAC ACCATGGCCG
2101 GGTGGGAATC TTCCCACGCA CCTATATCGA GCTTCTTCCT CCAGCTGAGA
2151 AGGCTCAGCC CAGAAAGTTG GCACCCGTAC AAGTTTTGGA ATATGGAGAA
```

FIG. 42A

2201 GCCATTGCAA AGTTTAACTT TAATGGAGAT ACACAAGTAG AAATGTCTTT
2251 CCGAAAGGGG GAGAGGATCA CGCTGCTCCG ACAGGTGGAT GAGAACTGGT
2301 ATGAAGGGAG GATTCCTGGG ACATCTCGCC AAGGCATTTT CCCTATCACC
2351 TATGTAGATG TGCTTAAGAG GCCATTGGTG AAAACCCCTG TGGATTACAT
2401 CGACCTGCCT TATTCTTCTT CCCCAAGTCG CAGTGCCACT GTGAGCCCAC
2451 AGGCTTCTCA TCATTCATTG AGCGCAGGAC CTGATCTCAC AGAATCTGAA
2501 AAGAACTATG TGCAACCTCA AGCCCAGCAG CGAAGAGTCA CCCCAGACAG
2551 GAGTCAGCCC TCACTGGATT TGTGTAGCTA CCAAGCGTTA TATAGTTATG
2601 TGCCACAGAA CGATGATGAG TTGGAACTCC GAGATGGAGA TATTGTTGAT
2651 GTCATGGAAA AATGTGACGA TGGATGGTTT GTTGGCACTT CGAGAAGGAC
2701 GAGGCAGTTT GGTACTTTTC CAGGCAACTA TGTAAAACCT TTATATCTAT
2751 AAGAAGACTA AAAAGCACAG AGATTATTTT TTATCGGAGG ATGAAGCATC
2801 ATTCATGAAC TGGTCTCTTT ATTTAAGTAC TGAGTCAGTA AGAAAACTAA
2851 TGCAGTTGGT AAAGAAAGAA TTCAAAGAAG GAACAGAGAA GTGTGTTTGA
2901 AACCCATTGT GTATCAGGGA TTAACTATCT GCTGAAGACA TCTGTATTTA
2951 CATGACTGCT TCTGGGAGCT GCTCTAGCCC CCGCTGCTTG GGGAATCTGA
3001 TCTGGAGCAT GTCCATGAGC AACATTAGCC AAAAAAAAAA GCTTGGGCCC
3051 TATTCTATAG TGTCACCTAA ATACTAGCTT GATCCGGCTG CTAACAAAGC
3101 CCGAAAGGAA GCTGAGTTGC TGCTGCCACC GCTGAGCAAT AACTAGCATA
3151 ACCCCTTGGG GCCTCTAAAC GGGTCTTGAG GGGTTTTTTG GCTGAAAGGA
3201 GGAACTATAT CCGGATAACC TGGCGTAATA GCGAAGAGGC CCGCACCGAT
3251 CGCCCTTCCC AACAGTTGGG CAGCCTGAAT GGCGAATGGA CGCGCCCTGT
3301 AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA GGGTG

FIG. 42B

```
   1 TTNNCACTCA CCGTCCTGGT GATGGTACCG GATCGAATTC AAGCGTGGCC
  51 GTGGCCGTGG GGCGCGCGGG GACCGCCCGG GGTGCCCGCT CCGCTCAGCG
 101 TCCGGGCCGC GTGGTCCGGC GGAGCCCCGA GACCACCCCC GGGCGGGGCG
 151 CCGCCGCGAT GTCGGTGGCT GGGCTCAAGA AGCAGTTCCA CAAAGCCAGC
 201 CAGCTGTTTA GTGAAAAAAT AAGTGGTGCC GAAGGAACGA AGCTAGATGA
 251 AGAATTTCTG AACATGGAAA AGAAAATAGA TATCACCAGT AAAGCTGTTG
 301 CAGAAATCCT TTCAAAAGCC ACAGAGTATC TCCAACCCAA TCCAGCATAC
 351 AGAGCTAAGC TAGGAATGCT GAACACTGTG TCGAAGCTCC GAGGGCAGGT
 401 GAAGGCCACC GGCTACCCAC AGACGGAAGG CTTGCTGGGG GACTGCATGC
 451 TGAAGTATGG CAAGGAGCTC GGAGAAGACT CTGCTTTTGG CAACTCGTTG
 501 GTAGATGTTG GTGAGGCCCT GAAACTCATG GCTGAGGTGA AAGACTCTCT
 551 GGATATTAAT GTGAAGCAAA CTTTTATTGA CCCACTGCAG CTACTGCAAG
 601 ACAAAGATTT AAAGGAGATC GGGCACCACC TGAGAAAGCT GGAAGGCCGT
 651 CGCCTGGATT ATGATTATAA AAAGCGGCGG GTAGGTAAGA TCCCCGAGGA
 701 AGAAATCAGA CAAGCAGTAG AGAAGTTTGA AGAGTCAAAG GAGTTGGCCG
 751 AAAGGAGCAT GTTTAATTTT TTAGAAAATG ATGTAGAGCA AGTGAGCCAG
 801 CTGGCTGTGT TTGTAGAGGC GGCATTAGAC TATCACAGGC AGTCCACAGA
 851 GATCCTCCAG GAGCTGCAGA GCAAGCTGGA GTTGCGAATA TCTCTTGCAT
 901 CCAAAGTCCC CAAGCGAGAA TTCATGCCAA AGCCTGTGAA CATGAGTTCC
 951 ACCGATGCCA ATGGGGTCGG ACCCAGCTCT TCATCAAAGA CACCAGGTAC
1001 TGACACTCCC GCGGACCAGC CCTGCTGTCG TGGTCTCTAT GACTTTGAGC
1051 CAGAAAATGA AGGAGAATTA GGATTTAAAG AAGGGGACAT CATTACATTA
1101 ACCAATCAGA TAGATGAAAA CTGGTATGAA GGGATGCTTC GTGGGGAATC
1151 CGGCTTCTTC CCCATTAATT ACGTGGAAGT CATTGTGCCT TTACCTCCGT
1201 AAATGTGTCT TTTGGACCTA ACTTCAGAAC TGAAATGAAT TGGCACCAGT
1251 GCTCTCTCAG TGTGGTGTTC TGTGACANCC TCGCTCTCTG GCCCACTTAA
1301 TCACTTTTGT ATGTGTGTTT TCTTTATAAT GTATTTTGTA TCAATTTAAT
1351 TTGTATAACT GATTTCTTTG TCCTAACTCA TAAAAATAGT TTTCTTCTTG
1401 TTCTAAAAAG TCATTGGTTA AATGTATTTG CTTCCTGTTG CTAAAACGAG
1451 TAAATTGCGC CCATTCGAAT GGCCTGGGTA GTCCTTGACT GCAGTGGGAA
1501 CGCACCCTTT GCAGCCATGA AAGCTAAAGG TTTGTTTCCT GACATTATTG
1551 ATGGCCTCTG GTCTTTTCCT GTTTTAAGCT TACCTGTGAA CAGCCCAATA
1601 AACNTGACAC ACTGTANAAT AANAAGGGTG GCCCNA
```

FIG. 44

```
1   MSVAGLKKQF HKASQLFSEK ISGAEGTKLD EEFLNMEKKI DITSKAVAEI
51  LSKATEYLQP NPAYRAKLGM LNTVSKLRGQ VKATGYPQTE GLLGDCMLKY
101 GKELGEDSAF GNSLVDVGEA LKLMAEVKDS LDINVKQTFI DPLQLLQDKD
151 LKEIGHHLRK LEGRRLDYDY KKRRVGKIPE EEIRQAVEKF EESKELAERS
201 MFNFLENDVE QVSQLAVFVE AALDYHRQST EILQELQSKL ELRISLASKV
251 PKREFMPKPV NMSSTDANGV GPSSSSKTPG TDTPADQPCC RGLYDFEPEN
301 EGELGFKEGD IITLTNQIDE NWYEGMLRGE SGFFPINYVE VIVPLPP
```

FIG. 45

```
1   MSGSYDEASE EITDSFWEVG NYKRTVKRID DGHRLCNDLM SCVQERAKIE
51  KAYAQQLTDW AKRWRQLIEK GPQYGSLERA WGAMMTEADK VSELHQEVKN
101 SLLNEDLEKV KNWQKDAYHK QIMGGFKETK EAEDGFRKAQ KPWAKKMKEL
151 EAAKKAYHLA CKEERLAMTR EMNSKTEQSV TPEQQKKLVD KVDKCRQDVQ
201 KTQEKYEKVL EDVGKTTPQY MEGMEQVFEQ CQQFEEKRLV FLKEVLLDIK
251 RHLNLAENSS YMHVYRELEQ AIRGADAQED LRWFRSTSGP GMPMNWPQFE
301 EWNPDLPHTT AKKEKQPKKA EGATLSNATG AVESTSQAGD RGSVSSYDRG
351 QTYATEWSDD ESGNPFGGNE ANGGANPFED DAKGVRVRAL YDYDGQEQDE
401 LSFKAGDELT KLGEEDEQGW CRGRLDSGQL GLYPANYVEA I
```

FIG. 47

```
   1 CGGGCTTGAG GCTGGGCCGC CGCCGCCGCC CGCTTTGCCA CCCGCCCCGC
  51 TGATGGTGTC CGGTGCTCCG GCGCCCAGGG ACACAGACCG GGAGCAGGAC
 101 CACTTCTCTC ACCTCCGGAT CTCTCCTGCT TCCGCAGCCT GTGAGCAGCA
 151 GGCCTGCTAA CTGCAGATCC ACAACCGCAC AGCTCGCTAC AGGTGCACCA
 201 TGTCTGGCTC CTACGATGAG GCCTCAGAGG AGATCACAGA TAGCTTCTGG
 251 GAGGTGGGGA ACTACAAGCG GACGGTGAAG CGCATCGACG ATGGGCACCG
 301 CCTGTGCAAC GACCTCATGA GCTGCGTGCA GGAGCGCGCC AAGATCGAGA
 351 AGGCATACGC GCAGCAGCTC ACCGACTGGG CCAAGCGCTG GCGCCAGCTC
 401 ATCGAGAAAG GTCCTCAGTA TGGCAGCCTG GAGCGGGCGT GGGGCGCCAT
 451 GATGACAGAA GCAGATAAGG TCAGCGAGCT GCACCAGGAG GTGAAGAACA
 501 GCCTGCTGAA TGAGGACCTG GAGAAAGTCA AGAACTGGCA GAAGGATGCC
 551 TATCACAAGC AGATCATGGG TGGCTTCAAG GAGACGAAAG AGGCCGAGGA
 601 TGGCTTCCGA AAGGCCCAGA AGCCCTGGGC TAAAAAGATG AAGGAGCTAG
 651 AGGCGGCCAA GAAGGCCTAT CACTTGGCTT GTAAGGAGGA AAGGCTGGCC
 701 ATGACCCGGG AGATGAACAG TAAGACAGAG CAGTCGGTCA CCCCTGAACA
 751 GCAGAAGAAA CTTGTGGACA AAGTGGACAA ATGCAGACAG GATGTGCAAA
 801 AGACTCAGGA GAAGTATGAG AAGGTCCTGG AAGATGTGGG CAAGACCACA
 851 CCACAGTACA TGGAGGGCAT GGAGCAGGTG TTTGAGCAGT GCCAGCAGTT
 901 TGAGGAGAAG CGGCTGGTCT TCCTGAAGGA AGTCCTGCTG GATATCAAAC
 951 GGCATCTCAA CCTAGCGGAG AACAGCAGCT ACATGCATGT CTACCGAGAA
1001 CTGGAGCAGG CCATCCGGGG GGCCGATGCC CAGGAGGACC TCAGGTGGTT
1051 CCGCAGCACC AGTGGCCCCG GGATGCCCAT GAACTGGCCG CAGTTCGAGG
1101 AGTGGAACCC AGACCTCCCG CACACCACTG CCAAGAAGGA GAAACAGCCT
1151 AAGAAGGCAG AGGGGGCCAC CCTGAGCAAT GCCACTGGGG CTGTAGAATC
1201 CACATCCCAG GCTGGGGACC GTGGCAGTGT TAGCAGCTAT GACCGAGGCC
1251 AAACATATGC CACCGAGTGG TCAGACGATG AGAGCGGAAA CCCCTTCGGG
1301 GGCAATGAGG CCAATGGTGG CGCCAACCCC TTCGAGGATG ATGCCAAGGG
1351 AGTTCGTGTA CGGGCACTCT ATGACTACGA CGGTCAGGAG CAGGATGAGC
1401 TCAGCTTCAA GGCCGGAGAT GAGCTCACCA AGCTCGGAGA GGAAGACGAA
1451 CAGGGTTGGT GCCGCGGGCG GCTGGACAGC GGACAGCTGG GCCTCTATCC
1501 TGCCAACTAC GTTGACGCTA TATAGCTACC TTGCCCACCC GACTCCTCTC
1551 AGTCCTTGTC CACCGCCTTC CACCCTTCCC CTCCCCCTTG CCATAGAGTT
1601 CCAGACATAT TTTCCCATCA AGCTTTTATT TTTTTAAAAG TCAAAACAGA
1651 ACAAAAAAAA AAAAAAAAAA GAAGAAATAC GAAGAGACAG CGTTTGCAGC
1701 CTACCTGGAG GCCGGGGGGG AGGGGGCTTA GGGTGATGGC CTCCCCCACA
1751 GCGTGGGCAA GGATCTTGGG ACTAACCCAA TGTCACATCT GGTCTATAGA
1801 GTCCACCAAA GAGTCTCCTG AGTCTTGAGG GAGATCTTCT GGATCCTTCT
1851 ACCCTGTCTC GCTCTCCTAT CCCACCACAG CTGCCAGCAG CTGCCCATGT
1901 CACCTGAGCC TGGCTTCCTA AACTCTCCTG TCCCCTCTCC TGTCCCCCTT
1951 CAACGCCCCC TTCTCTTAAA GGGCCCCCAA TCTTTAGTCT TCCACTCTGC
2001 CCTGGGGGTG CTTTTCTCTT CCCAGCCCTG TCCAGTGAGG CTGGGGGAGA
2051 AGGCTGCGGA GGGGAGGGGA GTGTCTCTTC ACTCCCCCAG ACATGAAGGC
2101 AGGTGAGTGG GAGGGAGTCA TGGCCTCCCT GGCATACAGG AGAGGAAGAA
2151 GGAGAACAGA CCATCTGACC AGGCTGTGCA ACACTCCCAA TGCCAAGCCC
```

FIG. 46A

2201 ATTTGAGGGA TGAAAACCCT AGCTGGGCCT GTGGGCAGAG GGCTCCTCCT
2251 CAGAGCCAAT GAGCATTTGC AGAGACCCTA CCTGTCTCTT TAGTCCTTGG
2301 CAATGGGCAA AGCCTCTTCC TTGGAAAGTC CAGGGCAAAG CCAGCAACAG
2351 TAGCAACCTC CTCTCACTCT GGGGAGGAGG CATTGGCCAC CCATCCCCCT
2401 CCCTTCATGG TCATTCAGAA ACGCCACAGC CCCTCCCATC CCCAATCACT
2451 GTGTCAGCAT CAGCCTTTGT GAAGACGGTC TACAAGGCTC TCACCTGGCC
2501 AACCTAGGAG ATTCAGGGGC TCAGGAACCT AGGAGATTCA GGGGCTTGGG
2551 GAACCTCCAC CTTGGCACTG TAAGGGGAAG CCAGCAGCTC AGGCTGGTGT
2601 GAGGAAGGAA CTCTGGATGG TCACTGTAGC TTTCTTCCTT GACCTTTTAG
2651 TCCCCAACAT CCCCTCTGAA TGCTGGCAGC ACCCCCACCC CCACACACAC
2701 ACTCCCATTT CTCTAAGCCC GAGAGTCTTG AGTCTTCATT AAAGGATTCT
2751 GGGTGTGGGA GGGGACACAG GGCCTTGTGG TTGGGAAGCA GGTGGCAGGC
2801 TCTCCCTTGG GAGGATGGGG TGGGAAACGA AACAGGTCAA CCAAGACCTC
2851 TTACAGTGGA AAGTGGTCAG AGGCTGTTTC TTTGGACCTT TGGGAACACA
2901 GATTTGAGAA AGTCTCATAT TCACAGCTGG TGTCCGCTAG GCCTCTGGCC
2951 TACGGACACC CTCTGCCTTG TGAATCAGGT GACCTTTTGG GCCTCCAGGG
3001 AAAGAACAGG ACCACCATCC ATGTTCTCCG CGTCCCTTTA GCTCTCTGCT
3051 GCTTCTCCTG ACACTCAGGT CATGGACCCA AGCTTTGGGG TCCTGACCAC
3101 CGCCCCCCCC CACCCCCCTT CTCTTGACTA GGCTGCAGCA GGGCCTTCTG
3151 TTGGGTCAGT CCTCCTCAGG GCCAGGAGCA GGAACTTAGC ACTCAAGAGA
3201 CAGGGCTGTA AGCACCCACT TCCCTGTCAC TGTTTGCCCT TGGGGCTTCA
3251 GCTGCAGCCC AGGTTGGGCC CTGGAGCCCT CAGAACCGGA AGCAGGATTC
3301 AAACCTCCCC TTCTCCACAG CCCCCCCTGC CTCCCCAGAT GGTAGACATC
3351 CCCCAGCTCT TACCTTCACC CTCATCTCAG AAAGGCAAGA AGCCGCCATG
3401 TCCGCACCTT GGGGCCTGGG CTTCCCCCTC TCTGTGCCAG CGGTTCCCAG
3451 CACCTGGGGA GGGGCTGTGG CCTGACCAGA CCCCAGGCCC ACCCCACATA
3501 GTATACTAGC TGCCCACTCT GGGGCAGGAA CTGGAAAATC CATCCCTTTT
3551 GAACAACCAC CTTCAATGAC CCCCCCCATC TGGGACCAGA CTTGGTCCTC
3601 AAGTTATTCA GCACCCCCAG TGCAGGAGGG TCCTCCCCCC ACCCCCCGAA
3651 GTCCCTGGAG CCCGGAGCAG AGCCCCACCT GTGATTCCTG GTGTTAGGGC
3701 ACCTCAAACC TTGGGCTGGA CCACACCCCT TCCCGCCATT TCCAGACCCC
3751 TACCTGTACT CCCCAGTGCT CCCCAGGGGC CTCTTGATGC TGCACGGGAC
3801 CCTGCAGGGC TCGGTCAGTG ATGTGTTTTG TCCCCAGTTA ACCGCCATCC
3851 AGCGACCTGG TTCCAGGAGG AGCTCAGGTC ACCCCCACCA CCGCCGCCAC
3901 TGCGTCTGCC GCCCTAGGCT TTCAGACATC ATTAGTTCCG ACACTTGTGA
3951 AACTCCGAGA CGTGCCGTGG TCTCAGCAAT GCACCTGTTT TATACATGAT
4001 TGTGTAATTT AAAGGTATAT AAATACAAAT ATATATATTA TATCTATATC
4051 TATCAGTTGT GACCGTATGG CTGTCGATAA AACCAGAATT C

FIG. 46B

```
1     GAATTCGTCG ACCCACGGTC CGGGAAGCCT TTCACAAGCA GATGATGGGC
51    GGCTTCAAGG AGACCAAGGA AGCTGAGGAC GGCTTTCGGA AGGCACAGAA
101   GCCCTGGGCC AAGAAGCTGA AAGAGGTAGA AGCAGCAAAG AAAGCCCACC
151   ATGCAGCGTG CAAAGAGGAG AAGCTGGCTA TCTCACGAGA AGCCAACAGC
201   AAGGCAGACC CATCCCTCAA CCCTGAACAG CTCAAGAAAT TGCAAGACAA
251   AATAGAAAAG TGCAAGCAAG ATGTTCTTAA GACCAAAGAG AAGTATGAGA
301   AGTCCCTGAA GGAACTCGAC CAGGGCACAC CCCAGTACAT GGAGAACATG
351   GAGCAGGTGT TTGAGCAGTG CCAGCAGTTC GAGGAGAAAC GCCTTCGCTT
401   CTTCCGGGAG GTTCTGCTGG AGGTTCAGAA GCACCTAGAC CTGTCCAATG
451   TGGCTGGTTA CAAAGCCATT TACCATGACC TGGAGCAGAG CATCAGAGCA
501   GCTGATGCAG TGGAGGACCT GAGGTGGTTC CGAGCCAATC ACGGGCCGGG
551   CATGGCCATG AACTGGCCGC AGTTTGAGGA GTGGTCCGCA GACCTGAATC
601   GAACCCTCAG CCGGAGAGAG AAGAAGAAGT CCACTGACGG CGTCACCCTG
651   ACGGGCATCA ACCAGACAGG CGACCAGTCT CTGCCGAGTA AGCCCAGCAG
701   CACCCTTAAT GTCCCGAGCA ACCCCGCCCA GTCTGCGCAG TCACAGTCCA
751   GCTACAACCC CTTCGAGGAT GAGGACGACA CGGGCAGCAC CGTCAGTGAG
801   AAGGACGACA CTAAGGCCAA AAATGTGAGC AGCTACGAGA AGACCCAGAG
851   CTATCCCACC GACTGGTCAG ACGATGAGTC TAACAACCCC TTCTCCTCCA
901   CGGATGCCAA TGGGGACTCG AATCCATTCG ACGACGACGC CACCTCGGGG
951   ACGGAAGTGC GAGTCCGGGC CCTGTATGAC TATGAGGGGC AGGAGCATGA
1001  TGAGCTGAGC TTCAAGGCTG GGGATGAGCT GACCAAGATG GAGGACGAGG
1051  ATGAGCAGGG CTGGTGCAAG GGACGCTTGG ACAACGGGCA AGTTGGCCTA
1101  TACCCGGCAA ATTATGTGGA GGCGATCCAG TGA
```

FIG. 48

```
1   RIRRPTVREA FHKQMMGGFK ETKEAEDGFR KAQKPWAKKL KEVEAAKKAH
51  HAACKEEKLA ISREANSKAD PSLNPEQLKK LQDKIEKCKQ DVLKTKEKYE
101 KSLKELDQGT PQYMENMEQV FEQCQQFEEK RLRFFREVLL EVQKHLDLSN
151 VAGYKAIYHD LEQSIRAADA VEDLRWFRAN HGPGMAMNWP QFEEWSADLN
201 RTLSRREKKK STDGVTLTGI NQTGDQSLPS KPSSTLNVPS NPAQSAQSQS
251 SYNPFEDEDD TGSTVSEKDD TKAKNVSSYE KTQSYPTDWS DDESNNPFSS
301 TDANGDSNPF DDDATSGTEV RVRALYDYEG QEHDELSFKA GDELTKMEDE
351 DEQGWCKGRL DNGQVGLYPA NYVEAIQ
```

FIG. 49

```
   1                AAAGGGAGG AGAGTGTCAA AAAGAAGGAT
  30 GGCGAGGAAA AAGGCAAACA GGAAGCACAA GACAAGCTGG
  70 GTCGGCTTTT CCATCAACAC CAAGAACCAG CTAAGCCAGC
 110 TGTCCAGGCA CCCTGGTCCA CTGCAGAAAA AGGGTCCACT
 150 TACCATTTCT GCACAGGAAA ATGTAAAAGT GGTGTATTAC
 190 CGGGCACTGT ACCCCTTTGA ATCCAGAAGC CATGATGAAA
 230 TCACTATCCA GCCAGGAGAC ATAGTCATGG TGGATGAAAG
 270 CCAAACTGGA GAACCCGGCT GGCTTGGAGG AGAATTAAAA
 310 GGAAAGACAG GGTGGTTCCC TGCAAACTAT GCAGAGAAAA
 350 TCCCAGAAAA TGAGGTTCCC GCTCCAGTGA AACCAGTGAC
 390 TGATTCAACA TCTGCCCCTG CCCCCAAACT GGCCTTGCGT
 430 GAGACCCCCG CCCCTTTGGC AGTAACCTCT TCAGAGCCCT
 470 CCACGACCCC TAATAACTGG GCCGACTTCA GCTCCACGTG
 510 GCCCACCAGC ACGAATGAGA AACCAGAAAC GGATAACTGG
 550 GATGCATGGG CAGCCCAGCC CTCTCTCACC GTTCCAAGTG
 590 CCGGCCAGTT AAGGCAGAGG TCCGCCTTTA CTCCAGCCAC
 630 GGCCACTGGC TCCTCCCCGT CTCCTGTGCT AGGCCAGGGT
 670 GAAAAGGTGG AGGGGCTACA AGCTCAAGCC CTATATCCTT
 710 GGAGAGCCAA AAAAGACAAC CACTTAAATT TTAACAAAAA
 750 TGATGTCATC ACCGTCCTGG AACAGCAAGA CATGTGGTGG
 790 TTTGGAGAAG TTCAAGGTCA GAAGGGTTGG TTCCCCAAGT
 830 CTTACGTGAA ACTCATTTCA GGGCCCATAA GGAAGTCTAC
 870 AAGCATGGAT TCTGGTTCTT CAGAGAGTCC TGCTAGTCTA
 910 AAGCGAGTAG CCTCTCCAGC AGCCAAGCCG GTCGTTTCGG
 950 GAGAAGAAAT TGCCCAGGTT ATTGCCTCAT ACACCGCCAC
 990 CGGCCCCGAG CAGCTCACTC TCGCCCCTGG TCAGCTGATT
1030 TTGATCCGAA AAAAGAACCC AGGTGGATGG TGGGAAGGAG
1070 AGCTGCAAGC ACGTGGGAAA AAGCGCCAGA TAGGCTGGTT
1110 CCCAGCTAAT TATGTAAAGC TTCTAAGCCC TGGGACGAGC
1150 AAAATCACTC CAACAGAGCC ACCTAAGTCA ACAGCATTAG
1190 CGGCAGTGTG CCAGGTGATT GGGATGTACG ACTACACCGC
1230 GCAGAATGAC GATGAGCTGG CCTTCAACAA GGGCCAGATC
1270 ATCAACGTCC TCAACAAGGA GGACCCTGAC TGGTGGAAAG
1310 GAGAAGTCAA TGGACAAGTG GGGCTCTTCC CATCCAATTA
1370 TGTGAAGCTG ACCACAGACA TGGACCCAAG CCAGCAATGA
```

FIG. 50

```
1   KGRRVSKRRM ARKKANRKHK TSWVGFSINT KNQLSQLSRH
41  PGPLQKKGPL TISAQENVKV VYYRALYPFE SRSHDEITIQ
81  PGDIVMVDES QTGEPGWLGG ELKGKTGWFP ANYAEKIPEN
121 EVPAPVKPVT DSTSAPAPKL ALRETPAPLA VTSSEPSTTP
161 NNWADFSSTW PTSTNEKPET DNWDAWAAQP SLTVPSAGQL
201 RQRSAFTPAT ATGSSPSPVL GQGEKVEGLQ AQALYPWRAK
241 KDNHLNFNKN DVITVLEQQD MWWFGEVQGQ KGWFPKSYVK
281 LISGPIRKST SMDSGSSESP ASLKRVASPA AKPVVSGEEI
321 AQVIASYTAT GPEQLTLAPG QLILIRKKNP GGWWEGELQA
361 RGKKRQIGWF PANYVKLLSP GTSKITPTEP PKSTALAAVC
401 QVIGMYDYTA QNDDELAFNK GQIINVLNKE DPDWWKGEVN
441 GQVGLFPSNY VKLTTDMDPS QQ
```

FIG. 51

```
1   GAATTCGCGG CCGCGTCGAC CAAGATCATT CCTGGGAGTG
41  AAGTAAAACG GGAAGAACCA GAAGCTTTGT ATGCAGCTGT
81  AAATAAGAAA CCTACCTCGG CAGCCTATTC AGTTGGAGAA
121 GAATATATTG CACTTTATCC ATATTCAAGT GTGGAACCTG
161 GAGATTTGAC TTTCACAGAA GGTGAAGAAA TATTGGTGAC
201 CCAGAAAGAT GGAGAGTGGT GGACAGGAAG TATTGGAGAT
241 AGAAGTGGAA TTTTTCCATC AAACTATGTC AAACCAAAGG
281 ATCAAGAGAG TTTTGGGAGT GCTAGCAAGT CTGGAGCATC
321 AAATAAAAAA CCTGAGATTG CTCAGGTAAC TTCAGCATAT
361 GTTGCTTCTG GTTCTGAACA ACTTAGCCTT GCACCAGGAC
401 AGTTAATATT AATTCTAAAG AAAAATACAA GTGGGTGGTG
441 GCAAGGAGAG TTACAGGCCA GAGGAAAAAA GCGACAGAAA
481 GGATGGTTTC CTGCCAGTCA TGTTAAACTT TTGGGTCCAA
521 GCAGTGAAAG AGCCACACCT GCCTTTCATC CTGTATGTCA
561 GGTGATTGCT ATGTATGACT ATGCAGCAAA TAATGAAGAT
601 GAGCTCAGTT TCTCCAAGGG ACAACTCATT AATGTTATGA
641 ACAAAGATGA TCCTGATTGG TGGCAAGGAG AGATCAACGG
681 GGTGACTGGT CTCTTTCCTT CAAACTACGT TAAGATGACG
721 ACAGACTCAG ATCCAAGTCA ACAGTGA
```

FIG. 52

```
1   EFAAASTKII PGSEVKREEP EALYAAVNKK PTSAAYSVGE
41  EYIALYPYSS VEPGDLTFTE GEEILVTQKD GEWWTGSIGD
81  RSGIFPSNYV KPKDQESFGS ASKSGASNKK PEIAQVTSAY
121 VASGSEQLSL APGQLILILK KNTSGWWQGE LQARGKKRQK
161 GWFPASHVKL LGPSSERATP AFHPVCQVIA MYDYAANNED
201 ELSFSKGQLI NVMNKDDPDW WQGEINGVTG LFPSNYVKMT
241 TDSDPSQQ
```

FIG. 53

```
HSLHLHRHQGRKERARYDLEAAQDNELTFKAGEIMTVLDDSDPNWWKGETHQGIGLFPSN 60
FVTADLTAEPEMIKTEKKTVQFSDDVQVETIEPEPEPAFIDEDKMDQLLQMLQSTDPSDD 120
QPDLPELLHLEAMCHQMGPLIDEKLEDIDRKHSELSELNVKVMEALSLYTKLMNEDPMYS 180
MYAKLQNQPYYMQSSGVSGSQVYAGPPPSGAYLVAGNAQMSHLQSYSLPPEQLSSLSQAV 240
VPPSANPALPSQQTQAAYPNRSPGDLMKPGDSECRGSAEDSQMRISPPYFPTGQQA 296
```

FIG. 55

```
IRGRVDQGEWPLPGRGTPGPSGLCVPEDQCRVRDLKGWLDSFWAKAEKEE 50
ENRRLEEKRWAEEAQRQLEQERRERELREAARREQRYQEQGGEASPQSRT 100
WEQQQEVVSRNRNEQESAVHPREIFKQKERAMSTTSISSPQPGKLRSPFL 150
QKQLTQPETHFGREPAAAISRPRADLPAEEPAPSTPPCLVQAEEEAVYEE 200
PPEQETFYEQPPLVQQQGAGSEHIDHHIQGQGLSGQGLCARALYDYQAAD 250
DTEISFDPENLITGIEVIDEGWWRGYGPDGHFGMFPANYVELIDEAEGTS 300
CPSPLRHGFLIAGRGGLGVDIQHSSRNRTPSEDEASGLPPAWQTQPVTPN 350
AAMAW 355
```

FIG. 57

```
GRVDIERKRLELMQKKKLEDEAARKAKQGKENLWKENLRKEEEEKQKRLQEEKTQEKIQE 60
EERKAEEKQRETASVLVNYRALYPFEARNHDEMSFNSGDIIQVDEKTVGEPGWLYGSFQG 120
NFGWFPCNYVEKMPSSENEKAVSPKKALLPPTVSLSATSTSSEPLSSNQPASVTDYQNVS 180
FSNLTVNTSWQKKSAFTRTVSPGSVSPIHGQGQVVENLKAQALCSWTAKKDNHLNFSKHD 240
IITVLEQQENWWFGEVHGGRGWFPKSYVKIIPGSEVKREEPEALYAAVNKKPTSAAYSVG 300
EEYIALYPYSSVEPGDLTFTEGEEILVTQKDGEWWTGSIGDRSGIFPSNYVKPKDQESFG 360
SASKSGASNKKPEIAQVTSAYVASGSEQLSLAPGQLILILKKNTSGWWQGELQARGKKRQ 420
KGWFPASHVKLLGPSSERATPAFHPVCQVIAMYDYAANNEDELSFSKGQLINVMNKDDPD 480
WWQGEINGVTGLFPSNYVKMTTDSDPSQQ 509
```

FIG. 59

CACTCTCTACACTTGCACCGGCATCAAGGACGAAAAGAAC 40
GCGCTAGATATGACTTGGAAGCTGCTCAAGACAATGAACT 80
TACTTTCAAAGCTGGAGAAATTATGACAGTTCTTGATGAC 120
AGTGATCCTAACTGGTGGAAAGGTGAAACCCATCAAGGCA 160
TAGGGTTATTTCCTTCTAATTTTGTGACTGCAGATCTCAC 200
TGCTGAACCAGAAATGATTAAAACAGAGAAGAAGACGGTA 240
CAATTTAGTGATGATGTTCAGGTAGAGACAATAGAACCAG 280
AGCCGGAACCAGCCTTTATTGATGAAGATAAAATGGACCA 320
GTTGCTACAGATGCTGCAAAGTACAGACCCCAGTGATGAT 360
CAGCCAGACCTACCAGAGCTGCTTCATCTTGAAGCAATGT 400
GTCACCAGATGGGACCTCTCATTGATGAAAAGCTGGAAGA 440
TATTGATAGAAAACATTCAGAACTCTCAGAACTTAATGTG 480
AAAGTGATGGAGGCCCTTTCCTTATATACCAAGTTAATGA 520
ACGAAGATCCGATGTATTCCATGTATGCAAAGTTACAGAA 560
TCAGCCATATTATATGCAGTCATCTGGTGTTTCTGGTTCT 600
CAGGTGTATGCAGGGCCTCCTCCAAGTGGTGCCTACCTGG 640
TTGCAGGGAACGCGCAGATGAGCCACCTCCAGAGCTACAG 680
TCTTCCCCCGGAGCAGCTGTCTTCTCTCAGCCAGGCAGTG 720
GTCCCACCATCCGCAAACCCAGCCCTTCCTAGTCAGCAGA 760
CTCAGGCCGCTTACCCAAACCGCTCCCCAGGGGACCTCAT 800
GAAGCCCGGTGATTCTGAATGCCGTGGATCTGCCGAGGAT 840
TCCCAGATGCGTATTTCTCCTCCGTACTTCCCCACAGGAC 880
AGCAGGCTTGAATAGCTGATTGCCTATGCAGGACAACAGG 920
CTTGAATAGCTGACTGCCTATGCATTCTCTTTGCTTGCCA 960
GTTTTTTGGACATCAAACTTGACAGATCCAAGATTATTAC 1000
TTTGATCTTCCCCACACCCCTCCCACCCCCGAGTCTACTA 1040
TGGTCCCATCATAGTATTCTGAAAATCAGTGAATGGCCAC 1080
TCTACCAGTTATTTCTACCAGTTTTTAGGTTCTAAACCTC 1120
AGGCATTCTGGACTCTTCTGTTCATTATCATATTTTGAAG 1160
GCATTATCTTCAAAATCTATCTAGACTCTGACCCTTTCTC 1200
CCATCTCCACCATTACTGCCGTGGCTCTTCTGCTGGTCGG 1240
CTCTCTCCTGGTGGATCCGTAATAACCTGCAGTCAGCTAT 1280
CCTGGTCCAGAAGGGAACCCCGTTAAACCCTGTTGGAATC 1320
TTATCACGCTTCTGCTCCAGAACGAACCCAGTCTGTCTGT 1360
CTCACTCAGAGTGTAAGCTACAGTCCTTATTGTGGCCATC 1400
AGGTGCTGTGTGTTCTCCAGCCCCCTCCCCACCACCGCAG 1440
TCCTGCCGGTGATCTTAGCTGCTCTCCCCTCGGAACCCCC 1480
TGCGGCCCCCTCTGCCGCAACAXTCGTGGCCTGCTGTTCC 1520
TTGAACATGCTTGGTGTTTTCTCTCCTCAAAGGCTTCTTT 1560
CTGTTTACCTGAAATGTACTTGCCTAGGGAAATCTTATCC 1600
TGGCTCACTCCGCTTACTTTTTTCCACATCTTTGCTTAAA 1640
GTTATTGCCCTTATTGGAGAAGGCACCCCTACCATAAACT 1680
AGAAATCCCTTGCCCCCAAGCTGCTCCTTT 1710

FIG. 54

GAATTCGCGGCCGCGTCGACCAAGGAGAGTGGCCGCTTCC 40
AGGACGTGGGACCCCAGGCCCCAGTGGGCTCTGTGTACCA 80
GAAGACCAATGCCGTGTCAGAGATTTAAAGGGTTGGTTAG 120
ACAGCTTCTGGGCCAAAGCAGAGAAGGAGGAGGAGAACCG 160
TCGGCTGGAGGAAAAGCGGTGGGCCGAGGAGGCACAGCGG 200
CAGCTGGAGCAGGAGCGCCGGGAGCGTGAGCTGCGTGAGG 240
CTGCACGCCGGGAGCAGCGCTATCAGGAGCAGGGTGGCGA 280
GGCCAGCCCCCAGAGCAGGACGTGGGAGCAGCAGCAAGAA 320
GTGGTTTCAAGGAACCGAAATGAGCAGGAGTCTGCCGTGC 360
ACCCGAGGGAGATTTTCAAGCAGAAGGAGAGGGCCATGTC 400
CACCACCTCCATCTCCAGTCCTCAGCCTGGCAAGCTGAGG 440
AGCCCCTTCCTGCAGAAGCAGCTCACCCAACCAGAGACCC 480
ACTTTGGCAGAGAGCCAGCTGCTGCCATCTCAAGGCCCAG 520
GGCAGATCTCCCTGCTGAGGAGCCGGCGCCCAGCACTCCT 560
CCATGTCTGGTGCAGGCAGAAGAGGAGGCTGTGTATGAGG 600
AACCTCCAGAGCAGGAGACCTTCTACGAGCAGCCCCCACT 640
GGTGCAGCAGCAAGGTGCTGGCTCTGAGCACATTGACCAC 680
CACATTCAGGGCCAGGGGCTCAGTGGGCAAGGGCTCTGTG 720
CCCGTGCCCTGTACGACTACCAGGCAGCCGACGACACAGA 760
GATCTCCTTTGACCCCGAGAACCTCATCACGGGCATCGAG 800
GTGATCGACGAAGGCTGGTGGCGTGGCTATGGGCCGGATG 840
GCCATTTTGGCATGTTCCCTGCCAATTACGTGGAGCTCAT 880
TGATGAGGCTGAGGGCACATCTTGCCCTTCCCCTCTCAGA 920
CATGGCTTCCTTATTGCTGGAAGAGGAGGCCTGGGAGTTG 960
ACATTCAGCACTCTTCCAGGAATAGGACCCCCAGTGAGGA 1000
TGAGGCCTCAGGGCTCCCTCCGGCTTGGCAGACTCAGCCT 1040
GTCACCCCAAATGCAGCAATGGCCTGGTGATTCCCACACA 1080
TCCTTCCTGCATCCCCCGACCCTCCCAGACAGCTTGGCTC 1120
TTGCCCCTGACAGGATACTGAGCCAAGCCCTGCCTGTGGC 1160
CAAGCCCTGAGTGGCCACTGCCAAGCTGCGGGGAAGGGTC 1200
CTGAGCAGGGGCATCTGGGAGGCTCTGGCTGCCTTCTGCA 1240
TTTATTTGCCTTTTTTCTTTTTCTCTTGCTTCTAAGGGGT 1280
GGTGGCCACCACTGTTTAGAATGACCCTTGGGAACAGTGA 1320
ACGTAGAGAATTGTTTTTAGCAGAGTTTGTGACCAAAGTC 1360
AGAGTGGATCATGGTGGTTTGGCAGCAGGGAATTTGTCTT 1400
GTTGGAGCCTGCTCTGTGCTCCCCACTCCATTTCTCTGTC 1440
CCTCTGCCTGGGCTATGGGAAGTGGGGATGCAGATGGCCA 1480
AGCTCCCACCCTGGGTATTCAAAAACGGCAGACACAACAT 1520
GTTCCTCCACGCGGCTCACTCGATGCCTGCAGGCCCCAGT 1560
GTGTGCCTCAACTGATTCTGACTTCAGGAAAAGTAACACA 1600
GAGTGGCCTTGGCCTGTTGTCTTCCCCTATTTTCTGTCCC 1640
AGCTCATCCGTGGTCGAAGCGCCCGCGAATTCCAGCTGAG 1680
CGGCCGC 1687

FIG. 56

```
GCGGCCGCGTCGACATTGAAAGGAAAAGATTAGAACTAAT    40
GCAGAAAAAGAAACTAGAAGATGAGGCTGCAAGGAAAGCA    80
AAGCAAGGAAAAGAAAACTTATGGAAAGAAAATCTTAGAA   120
AGGAGGAAGAAGAAAAACAAAAGCGACTCCAGGAAGAAAA   160
AACACAAGAAAAAATTCAAGAAGAGGAACGGAAAGCTGAG   200
GAGAAACAACGTGAGACAGCTAGTGTTTTGGTGAATTATA   240
GAGCATTATACCCCTTTGAAGCAAGGAACCATGATGAGAT   280
GAGTTTTAATTCTGGAGATATAATTCAGGTTGATGAAAAA   320
ACCGTAGGAGAACCTGGTTGGCTTTATGGTAGTTTTCAAG   360
GAAATTTTGGCTGGTTTCCATGCAATTATGTAGAAAAAAT   400
GCCATCAAGTGAAAATGAAAAAGCTGTATCTCCAAAGAAG   440
GCCTTACTTCCTCCTACAGTTTCTTTATCTGCTACCTCAA   480
CTTCCTCTGAACCACTTTCTTCAAATCAACCAGCATCAGT   520
GACTGATTATCAAAATGTATCTTTTTCAAACCTAACTGTA   560
AATACATCATGGCAGAAAAAATCAGCCTTCACTCGAACTG   600
TGTCCCCTGGATCTGTATCACCTATTCATGGACAGGGACA   640
AGTGGTAGAAAACTTAAAAGCACAGGCCCTTTGTTCCTGG   680
ACTGCAAAGAAAGATAACCACTTGAACTTCTCAAAACATG   720
ACATTATTACTGTCTTGGAGCAGCAAGAAAATTGGTGGTT   760
TGGGGAGGTGCATGGAGGAAGAGGATGGTTTCCCAAATCT   800
TATGTCAAGATCATTCCTGGGAGTGAAGTAAAACGGGAAG   840
AACCAGAAGCTTTGTATGCAGCTGTAAATAAGAAACCTAC   880
CTCGGCAGCCTATTCAGTTGGAGAAGAATATATTGCACTT   920
TATCCATATTCAAGTGTGGAACCTGGAGATTTGACTTTCA   960
CAGAAGGTGAAGAAATATTGGTGACCCAGAAAGATGGAGA  1000
GTGGTGGACAGGAAGTATTGGAGATAGAAGTGGAATTTTT  1040
CCATCAAACTATGTCAAACCAAAGGATCAAGAGAGTTTTG  1080
GGAGTGCTAGCAAGTCTGGAGCATCAAATAAAAAACCTGA  1120
GATTGCTCAGGTAACTTCAGCATATGTTGCTTCTGGTTCT  1160
GAACAACTTAGCCTTGCACCAGGACAGTTAATATTAATTC  1200
TAAAGAAAAATACAAGTGGGTGGTGGCAAGGAGAGTTACA  1240
GGCCAGAGGAAAAAAGCGACAGAAAGGATGGTTTCCTGCC  1280
AGTCATGTTAAACTTTTGGGTCCAAGTAGTGAAAGAGCCA  1320
CACCTGCCTTTCATCCTGTATGTCAGGTGATTGCTATGTA  1360
TGACTATGCAGCAAATAATGAAGATGAGCTCAGTTTCTCC  1400
AAGGGACAACTCATTAATGTTATGAACAAAGATGATCCTG  1440
ATTGGTGGCAAGGAGAGATCAACGGGGTGACTGGTCTCTT  1480
TCCTTCAAACTACGTTAAGATGACGACAGACTCAGATCCA  1520
AGTCAACAGTGACCCAATGTTGTCTTCCAGTTGTGAAAGC  1560
ACCCCAGAGACCCACTATCCAAGTTTCACTCTAGCGTGGA  1600
GGCAGGGCAGGCAGCCCTGATCAAATATCTGCTACACAAT  1640
TCGTTTACTTCGTTTGAATGTTAGAGCCACTTGTGATTAT  1680
TTTTTTGTGTTTCTAACTTACAGTTTAAATTTATTTGTAA  1720
```

FIG. 58A

AAAGTTAAAGGATAGTGGGTCTTTGTGTGGCTTTCCCTGC 1760
TGTTCACTCTGGCATCTTTAGCATTTTTCTTCTTTTTTAA 1800
TTTGATAATTGTAGGTCATTAGCATGCATATTGAGTTTGC 1840
CCTTATGTGGTGGGAGTTCAAACACACAAAGACCCACTAT 1880
TTGCACAAACTATTCTTACTGGTTTGGAATAGGCTGCCAT 1920
GCTTTTTTAATGTTATTGCAACATGTGTATTCATTTACAG 1960
AATTCAGATAAAATTTGCTTATGTTCTGCTATTATGTTTG 2000
ATCTAATCCTAATCACAGTGAGCTCTTAATTAGCTCAATA 2040
TGTGGTTTGCCCTCAAGTGTGCACTGTTTATTACTTTGTA 2080
ATATGCCACTATGAGTACTGACATTTAGATATGTTTAAAG 2120
GCCAAGAACTGGAAACAGCCATGCCCTGTTTTCTGTGTAT 2160
TTGGGATGGGAATAACAACATTTTGGGGGGAGCTTTTTAA 2200
ATCTCAGAGAAGAGGAAAGTGGCCTGCTCTGGCAGGTATG 2240
TGCAGTGTTTCATTTGTTCCAGTCCCAAGAATGAGCACTG 2280
TCCTATGGTAGTTCGCTTAGGATCTTTATGTGCTCTGGGC 2320
TAATGAAGGTACTGCATCATGTGCTGCAGCGTGTGTATTC 2360
TTTTTCGATGACCTATAAAGGGATTATTTTTGAGGAATGA 2400
AAGGCTCCCATCATTGACTGTGAGATGGGAAAAACCTTTC 2440
CTAGCTTAGAGCATTTATATCTTAATCCATTTTAAAGTCA 2480
GAGTTCATTGTTACCTGTTTTAATCAGGTGACTACATGTC 2520
CCAGTATACAAAGGGGCACTGGTTGACATTCTTCTTAATG 2560
TATTTAGTAAATATCATAAGAAATCCTTTAAGAGTTTAAA 2600
TGTCCCCAAAACAGACATGCGGGCTCTAGTCAAGAATGAA 2640
TTAGAGTGAAGGAAAGCTGTGTAACACCTGGCATTCCTCT 2680
GTGTTCATGGAGCTTCTTTGAGGCTCTAAGATTGATTTTA 2720
CCATCAGACTTCTCTAATACCTGTTCTTCAACCATATTGG 2760
CTACTTTGACATAAGAATTTACTTCTTTTCCTGGAATGGA 2800
AAACACTTTAAAAAATAATAACAAACATTATTATAAACTA 2840
ATATATGTGAGAGGTCGACGCGGCCGCGAATTC 2873

FIG. 58B

GAATTCGTCGACCCACGCGTCCGAAATATAACTGAAGTTGGGGCACCTAC 50
TGAAGAAGAGGAAGAAAGTGAAAGTGAAGATAGTGAAGACAGTGGTGGGG 100
AGGAAGAAGATGCAGAGGAGGAAGAGGAAGAGAAAGAGGAAAATGAATCT 150
CACAAATGGTCAACCGGTGAAGAATACATCGCTGTTGGAGATTTTACTGC 200
TCAGCAAGTTGGAGATCTTACATTTAAGAAAGGGGAAATTCTCCTTGTAA 250
TTGAAAAAAAAACCTGATGGTTGGTGGATAGCTAAGGATGCCAAAGGAAAT 300
GAAGGTCTTGTTCCCAGAACCTACCTAGAGCCTTATAGTGAAGAAGAAGA 350
AGGCCAAGAGTCAAGTGAAGAGGGCAGTGAAGAAGATGTAGAGGCGGTGG 400
ATGAAACAGCAGATGGAGCAGAAGTTAAGCAAAGAACTGATCCCCACTGG 450
AGTGCTGTTCAGAAAGCGATTTCAGAGGCGGGCATCTTCTGTCTTGTTAA 500
TCATGTCTCGTTTTGCTACCTAATAGTTCTGATCCGTCCCTAA 543

FIG. 60

GAATTCGGCGGACTTCGCGGCCGCGTCGACGAAGAAACCT 40
GAAGGACACACTAGGCCTCGGCAAGACGCGCAGGAAGACC 80
AGCGCGCGGGATGCGTCCCCCACGCCCAGCACGGACGCCG 120
AGTACCCCGCCAATGGCAGCGGCGCCGACCGCATCTACGA 160
CCTCAACATCCCGGCCTTCGTCAAGTTCGCCTATGTGGCC 200
GAGCGGGAGGATGAGTTGTCCCTGGTGAAGGGGTCGCGCG 240
TCACCGTCATGGAGAAGTGCAGCGACGGTTGGTGGCGGGG 280
CAGCTACAACGGGCAGATCGGCTGGTTCCCCTCCAACTAC 320
GTCTTGGAGGAGGTGGACGAGGCGGTTGCGGAGTCCCCAA 360
GCTTCCTGAGCCTGCGCAAGGGCGCCTCGCTGAGCAATGG 400
CCAGGGCTCCCGCGTGCTGCATGTGGTCCAGACGCTGTAC 440
CCCTTCAGCTCAGTCACCGAGGAGGAGCTCAACTTCGAGA 480
AGGGGGAGACCATGGAGGTGATTGAGAAGCCGGAGAACGA 520
CCCCGAGTGGTGGAAATGCAAAAATGCCCGGGGCCAGGTG 560
GGCCTCGTCCCCAAAAACTACGTGGTGGTCCTCAGTGACG 600
GGCCTGCCCTGCACCCTGCGCACGCCCCACAGATAAGCTA 640
CACCGGGCCCTCGTCCAGCGGGCGCTTCGCGGGCAGAGAG 680
TGGTACTACGGGAACGTGACGCGGCACCAGGCCGAGTGCG 720
CCCTCAACGAGCGGGGCGTGGAGGGCGACTTCCTCATTAG 760
GGACAGCGAGTCCTCGCCCAGCGACTTCTCCGTGTCCCTT 800
AAAGCGTCAGGGAAGAACAAACACTTCAAGGTGCAGCTCG 840
TGGACAATGTCTACTGCATTGGGCAGCGGCGCTTCCACAC 880
CATGGACGAGCTGGTGGAACACTACAAAAAGGCGCCCATC 920
TTCACCAGCGAGCACGGGGAGAAGCTCTACCTCGTCAGGG 960
CCCTGCAGTGA 971

FIG. 62

```
GAATTCGTCGACCCACGCGTCCGAAATATAACTGAAGTTGGGGCACCTACTGAAGAAGAGGAAGAAAGTG
                                                                        70
 E  F  V  D  P  R  V  R  N  I  T  E  V  G  A  P  T  E  E  E  E  E  S     23
AAAGTGAAGATAGTGAAGACAGTGGTGGGGAGGAAGAAGATGCAGAGGAGGAAGAGGAAGAGAAAGAGGA
                                                                       140
 E  S  E  D  S  E  D  S  G  G  E  E  E  D  A  E  E  E  E  E  E  K  E  E  47
AAATGAATCTCACAAATGGTCAACCGGTGAAGAATACATCGCTGTTGGAGATTTTACTGCTCAGCAAGTT
                                                                       210
                                                    SH3
 N  E  S  H  K  W  S  T  G  E  E  Y  I  A  V [G  D  F  T  A  Q  Q  V     70
GGAGATCTTACATTTAAGAAAGGGGAAATTCTCCTTGTAATTGAAAAAAAACCTGATGGTTGGTGGATAG
                                                                       280
 G  D  L  T  F  K  K  G  E  I  L  L  V  I  E  K  K  P  D  G  W  W  I     93
CTAAGGATGCCAAAGGAAATGAAGGTCTTGTTCCCAGAACCTACCTAGAGCCTTATAGTGAAGAAGAAGA
                                                                       350
 A  K  D  A  K  G  N  E  G  L  V  P  R  T  Y] L  E  P  Y  S  E  E  E  E  117
AGGCCAAGAGTCAAGTGAAGAGGGCAGTGAAGAAGATGTAGAGGCGGTGGATGAAACAGCAGATGGAGCA
                                                                       420
 G  Q  E  S  S  E  E  G  S  E  E  D  V  E  A  V  D  E  T  A  D  G  A  140
GAAGTTAAGCAAAGAACTGATCCCCACTGGAGTGCTGTTCAGAAAGCGATTTCAGAGGCGGGCATCTTTT
                                                                       490
 E  V  K  Q  R  T  D  P  H  W  S  A  V  Q  K  A  I  S  E  A  G  I  F  163
```

FIG.61A

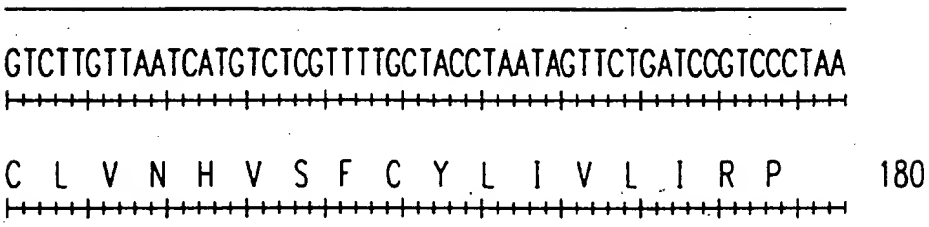
GTCTTGTTAATCATGTCTCGTTTTGCTACCTAATAGTTCTGATCCGTCCCTAA
C L V N H V S F C Y L I V L I R P
180

FIG.61B

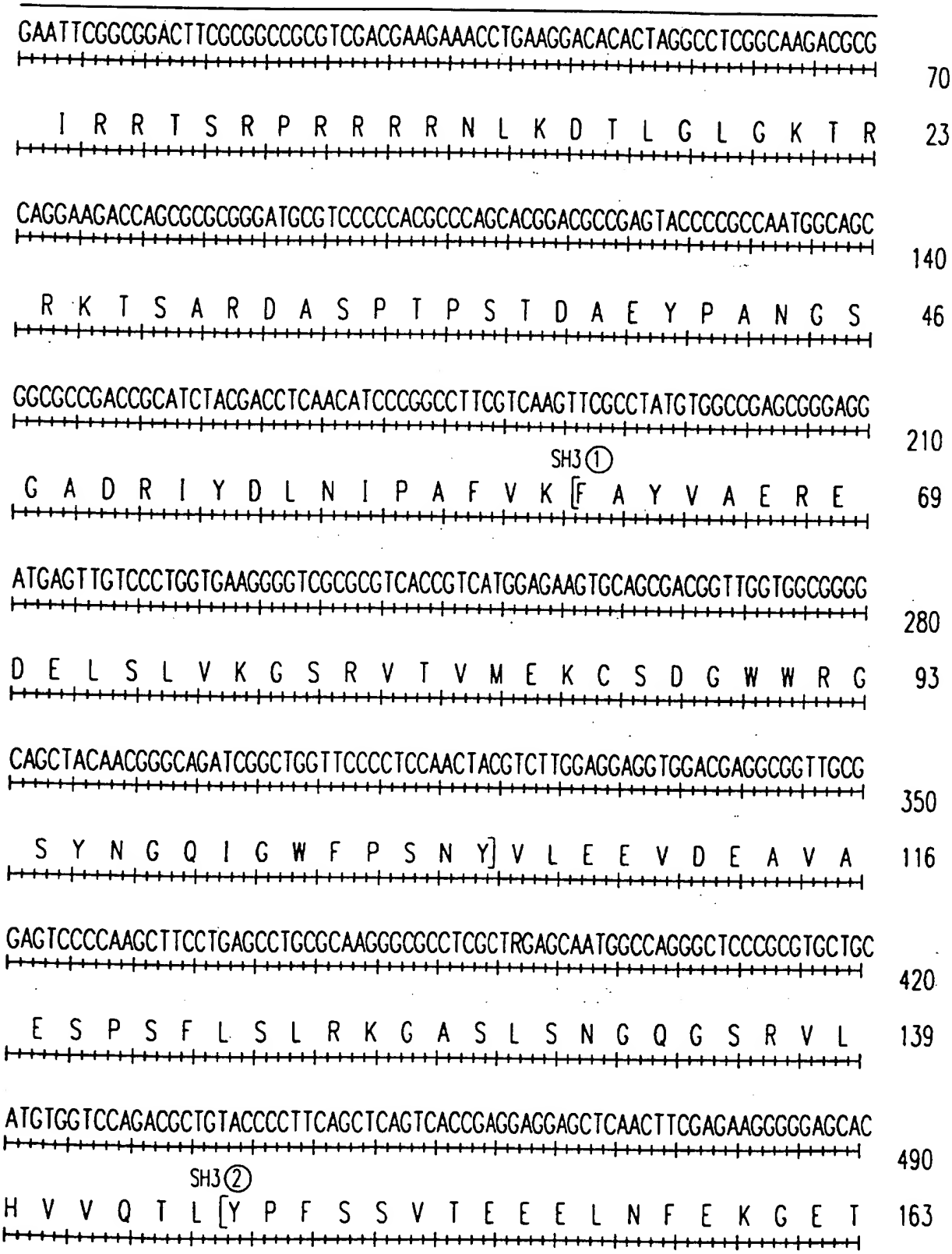
GAATTCGGCGGACTTCGCGGCCGCGTCGACGAAGAAACCTGAAGGACACACTAGGCCTCGGCAAGACGCG
70
I R R T S R P R R R R N L K D T L G L G K T R
23
CAGGAAGACCAGCGCGCGGGATGCGTCCCCCACGCCCAGCACGGACGCCGAGTACCCCGCCAATGGCAGC
140
R K T S A R D A S P T P S T D A E Y P A N G S
46
GGCGCCGACCGCATCTACGACCTCAACATCCCGGCCTTCGTCAAGTTCGCCTATGTGGCCGAGCGGGAGG
210
SH3①
G A D R I Y D L N I P A F V K [F A Y V A E R E
69
ATGAGTTGTCCCTGGTGAAGGGGTCGCGCGTCACCGTCATGGAGAAGTGCAGCGACGGTTGGTGGCGGGG
280
D E L S L V K G S R V T V M E K C S D G W W R G
93
CAGCTACAACGGGCAGATCGGCTGGTTCCCCTCCAACTACGTCTTGGAGGAGGTGGACGAGGCGGTTGCG
350
S Y N G Q I G W F P S N Y] V L E E V D E A V A
116
GAGTCCCCAAGCTTCCTGAGCCTGCGCAAGGGCGCCTCGCTRGAGCAATGGCCAGGGCTCCCGCGTGCTGC
420
E S P S F L S L R K G A S L S N G Q G S R V L
139
ATGTGGTCCAGACGCTGTACCCCTTCAGCTCAGTCACCGAGGAGGAGCTCAACTTCGAGAAGGGGGAGCAC
490
SH3②
H V V Q T L [Y P F S S V T E E E L N F E K G E T
163

FIG.63A

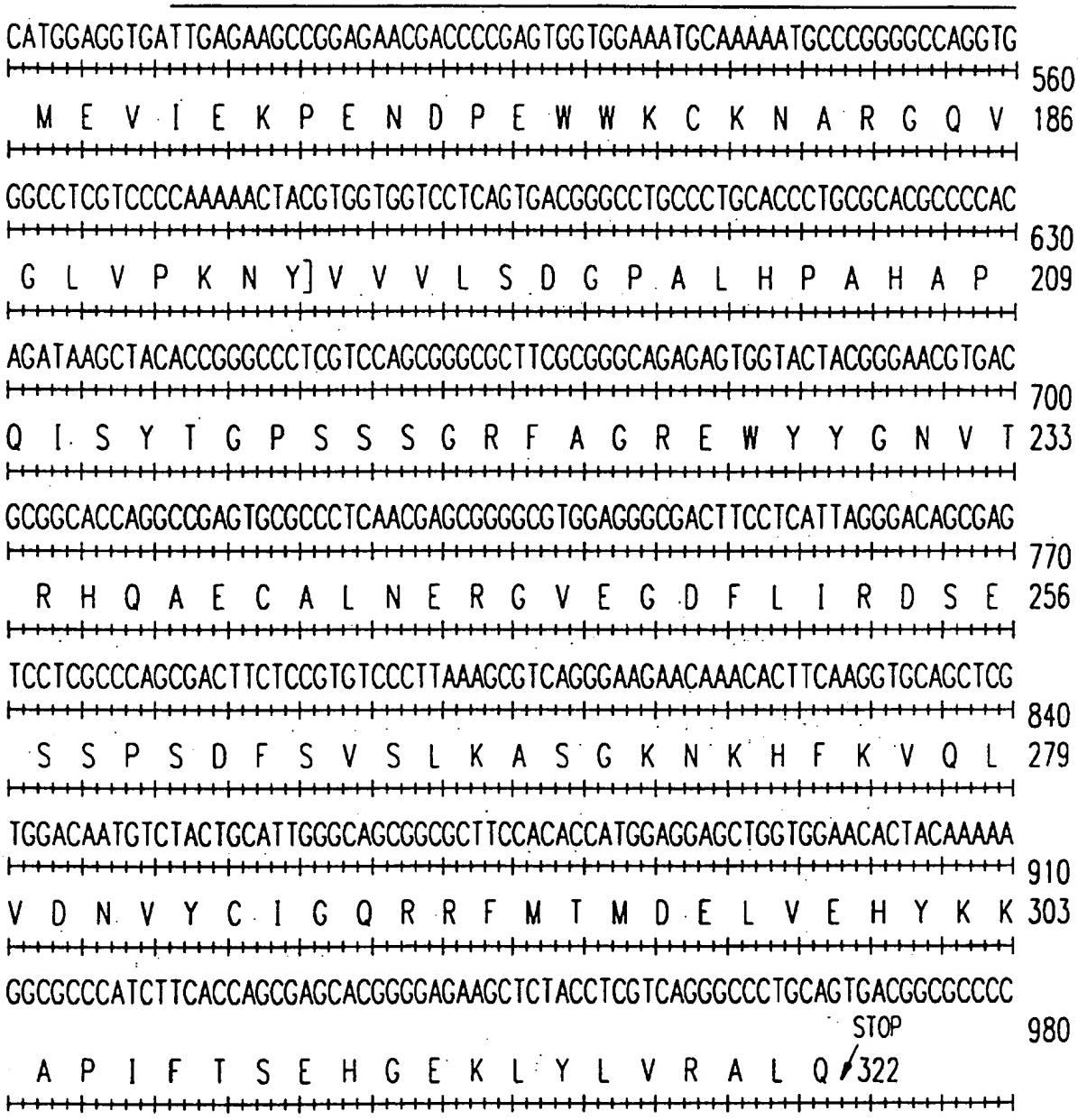
CATGGAGGTGATTGAGAAGCCGGAGAACGACCCCGAGTGGTGGAAATGCAAAAATGCCCGGGGCCAGGTG
560
M E V I E K P E N D P E W W K C K N A R G Q V 186
GGCCTCGTCCCCAAAAACTACGTGGTGGTCCTCAGTGACGGGCCTGCCCTGCACCCTGCGCACGCCCCAC
630
G L V P K N Y] V V V L S D G P A L H P A H A P 209
AGATAAGCTACACCGGGCCCTCGTCCAGCGGGCGCTTCGCGGGCAGAGAGTGGTACTACGGGAACGTGAC
700
Q I S Y T G P S S S G R F A G R E W Y Y G N V T 233
GCGGCACCAGGCCGAGTGCGCCCTCAACGAGCGGGGCGTGGAGGGCGACTTCCTCATTAGGGACAGCGAG
770
R H Q A E C A L N E R G V E G D F L I R D S E 256
TCCTCGCCCAGCGACTTCTCCGTGTCCCTTAAAGCGTCAGGGAAGAACAAACACTTCAAGGTGCAGCTCG
840
S S P S D F S V S L K A S G K N K H F K V Q L 279
TGGACAATGTCTACTGCATTGGGCAGCGGCGCTTCCACACCATGGAGGAGCTGGTGGAACACTACAAAAA
910
V D N V Y C I G Q R R F M T M D E L V E H Y K K 303
GGCGCCCATCTTCACCAGCGAGCACGGGGAGAAGCTCTACCTCGTCAGGGCCCTGCAGTGACGGCGCCCC
STOP
980
A P I F T S E H G E K L Y L V R A L Q /322

FIG.63B

GAATTCGCGGACTTCGCGGCCGCGTCGACACCAGTGCAGG 40
TTTTGGAATATGGAGAAGCTATTGCTAAGTTTAACTTTAA 80
TGGTGATACACAAGTAGAAATGTCCTTCAGAAAGGGTGAG 120
AGGATCACACTGCTCCGGCAGGTAGATGAGAACTGGTACG 160
AAGGGAGGATCCCGGGGACATCCCGACAAGGCATCTTCCC 200
CATCACCTACGTGGATGTGATCAAGCGACCACTGGTGAAA 240
AACCCTGTGGATTACATGGACCTGCCTTTCTCCTCCTCCC 280
CAAGTCGCAGTGCCACTGCAAGCCCACAGCAACCTCAAGC 320
CCAGCAGCGAAGAGTCACCCCCGACAGGAGTCAAACCTCA 360
CAAGATTTATTTAGCTATCAAGCATTATATAGCTATATAC 400
CACAGAATGATGATGAGTTGGAACTCCGCGATGGAGATAT 440
CGTTGATGTCATGGAAAAATGTGACGATGGATGGTTTGTT 480
GGTACTTCAAGAAGGACAAAGCAGTTTGGTACTTTTCCAG 520
GCAACTATGTAAAACCTTTGTATCTATAA

FIG. 64

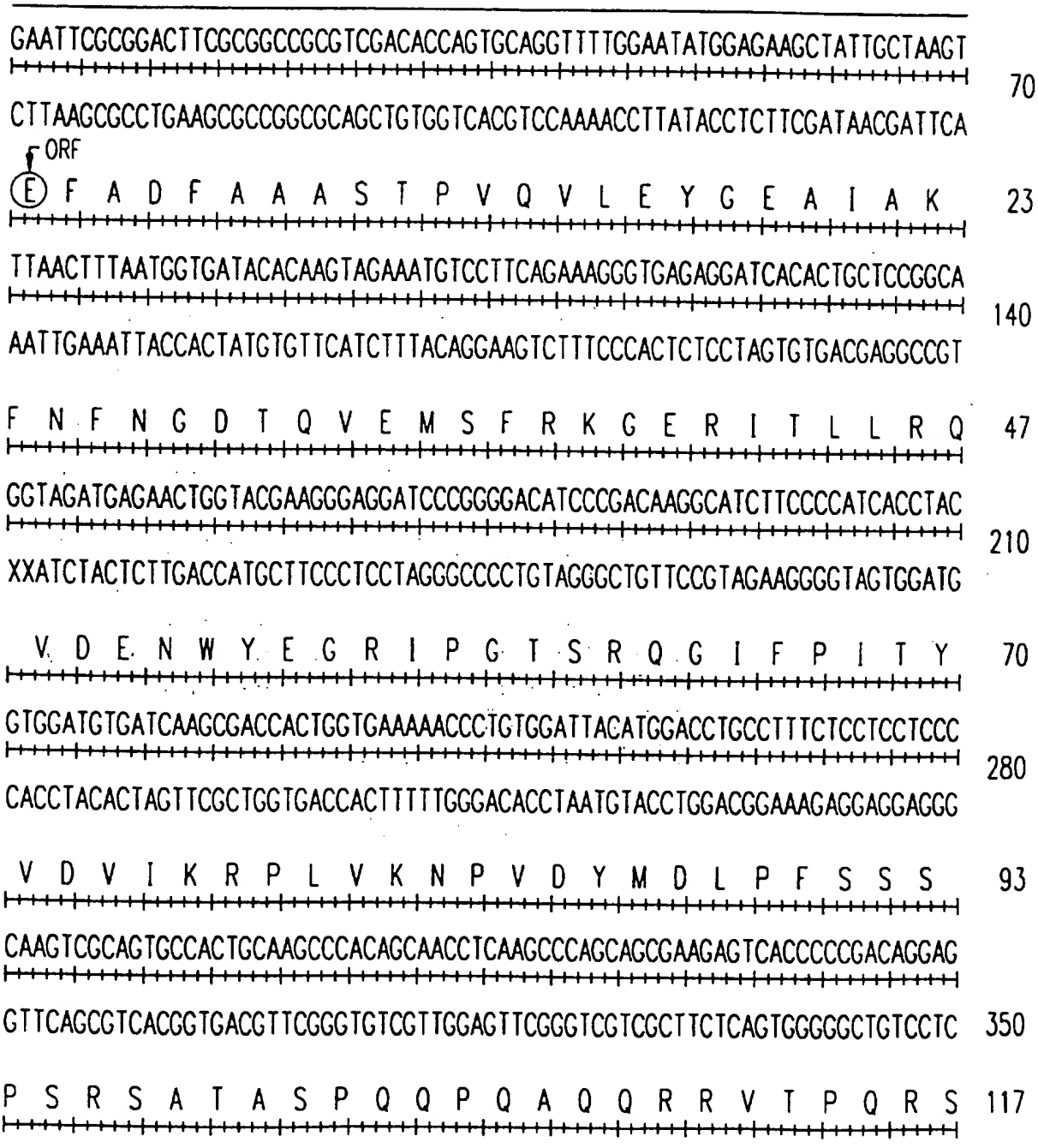
GAATTCGCGGACTTCGCGGCCGCGTCGACACCAGTGCAGGTTTTGGAATATGGAGAAGCTATTGCTAAGT
70
CTTAAGCGCCTGAAGCGCCGGCGCAGCTGTGGTCACGTCCAAAACCTTATACCTCTTCGATAACGATTCA
ORF
E F A D F A A A S T P V Q V L E Y G E A I A K
23
TTAACTTTAATGGTGATACACAAGTAGAAATGTCCTTCAGAAAGGGTGAGAGGATCACACTGCTCCGGCA
140
AATTGAAATTACCACTATGTGTTCATCTTTACAGGAAGTCTTTCCCACTCTCCTAGTGTGACGAGGCCGT
F N F N G D T Q V E M S F R K G E R I T L L R Q
47
GGTAGATGAGAACTGGTACGAAGGGAGGATCCCGGGGACATCCCGACAAGGCATCTTCCCCATCACCTAC
210
XXATCTACTCTTGACCATGCTTCCCTCCTAGGGCCCCTGTAGGGCTGTTCCGTAGAAGGGGTAGTGGATG
V D E N W Y E G R I P G T S R Q G I F P I T Y
70
GTGGATGTGATCAAGCGACCACTGGTGAAAAACCCTGTGGATTACATGGACCTGCCTTTCTCCTCCTCCC
280
CACCTACACTAGTTCGCTGGTGACCACTTTTTGGGACACCTAATGTACCTGGACGGAAAGAGGAGGAGGG
V D V I K R P L V K N P V D Y M D L P F S S S
93
CAAGTCGCAGTGCCACTGCAAGCCCACAGCAACCTCAAGCCCAGCAGCGAAGAGTCACCCCCGACAGGAG
GTTCAGCGTCACGGTGACGTTCGGGTGTCGTTGGAGTTCGGGTCGTCGCTTCTCAGTGGGGGCTGTCCTC
350
P S R S A T A S P Q Q P Q A Q Q R R V T P Q R S
117

FIG.65A

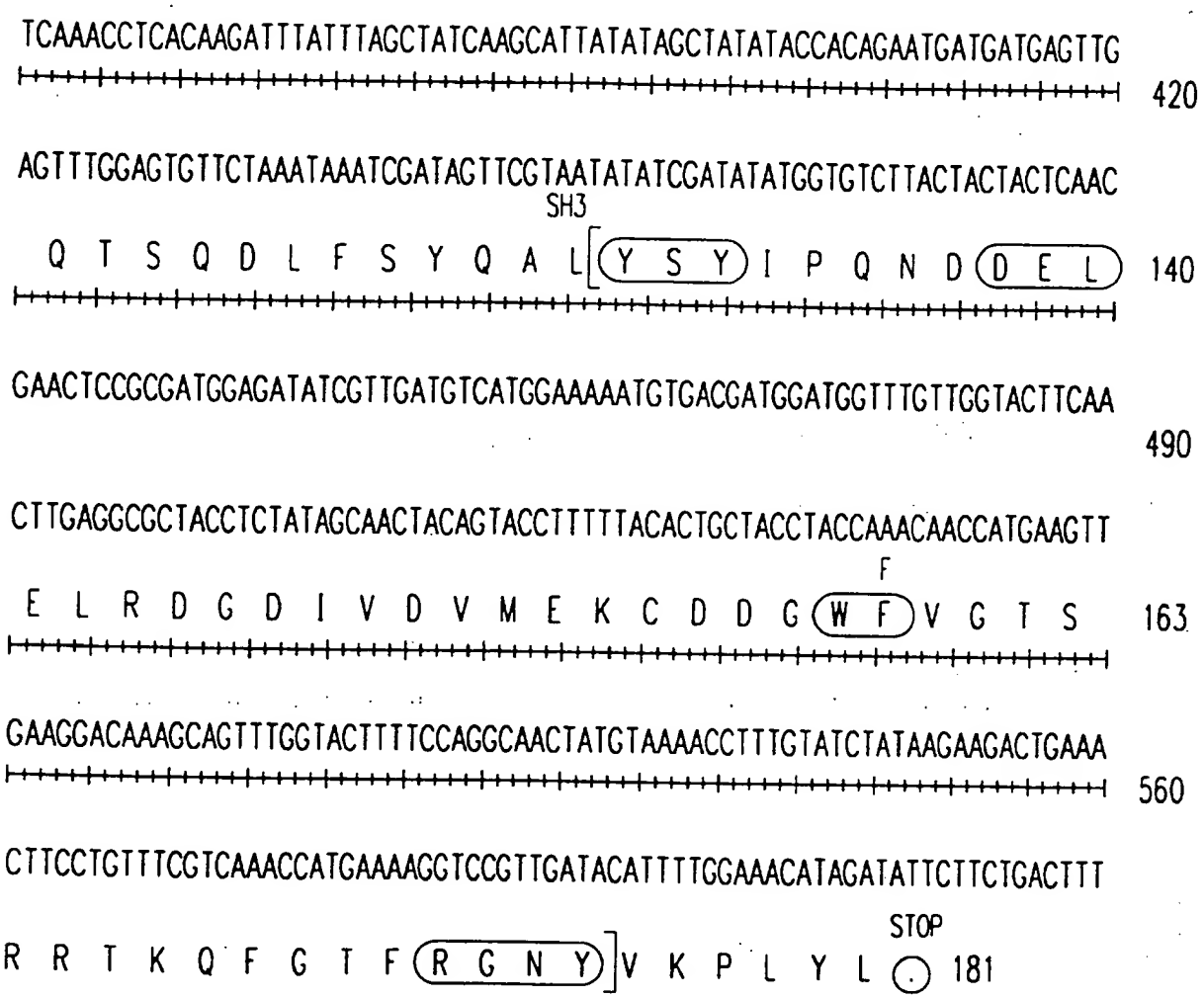
TCAAACCTCACAAGATTTATTTAGCTATCAAGCATTATATAGCTATATACCACAGAATGATGATGAGTTG
420
AGTTTGGAGTGTTCTAAATAAATCGATAGTTCGTAATATATCGATATATGGTGTCTTACTACTACTCAAC
SH3
Q T S Q D L F S Y Q A L [Y S Y] I P Q N D [D E L] 140
GAACTCCGCGATGGAGATATCGTTGATGTCATGGAAAAATGTGACGATGGATGGTTTGTTGGTACTTCAA
490
CTTGAGGCGCTACCTCTATAGCAACTACAGTACCTTTTTACACTGCTACCTACCAAACAACCATGAAGTT
F
E L R D G D I V D V M E K C D D G [W F] V G T S 163
GAAGGACAAAGCAGTTTGGTACTTTTCCAGGCAACTATGTAAAACCTTTGTATCTATAAGAAGACTGAAA
560
CTTCCTGTTTCGTCAAACCATGAAAAGGTCCGTTGATACATTTTGGAAACATAGATATTCTTCTGACTTT
STOP
R R T K Q F G T F [R G N Y] V K P L Y L ⊙ 181

FIG.65B

```
AATTCAAGCGCGGGGTCTTTAGGATTTGCAGCTCCAGGAAGCGAGATGTCGAAGCCGCCACCCAAACCA
                                                                      70
  N  S  S  A  G  S  L  G  F  A  A  P  G  S  E  M  S  K  P  P  P  K  P

GTCAAACCAGGGCAAGTTAAAGTCTTCAGAGCCCTGTATACGTTTGAACCCAGAACTCCAGATGAATTAT
                                                                     140
  V  K  P  G  Q  V  K  V  F  R  A  L  Y  T  F  E  P  R  T  P  D  E  L

ACTTTGAGGAAGGTGATATTATCTACATTACTGACATGAGCGATACCAATTGGTGGAAAGGCACCTCCAA
                                                                     210
  Y  F  E  E  G  D  I  I  Y  I  T  D  M  S  D  T  N  W  W  K  G  T  S  K

AGGCAGGACTGGACTAATTCCAAGCAACTATGTGGCTGAGCAGGCAGAATCCATTGACAATCCATTGCAT
                                                                     280
  G  R  T  G  L  I  P  S  N  Y  V  A  E  Q  A  E  S  I  D  N  P  L  H

GAAGCAGCAAAAAGAGGCAACTTGAGCTGGTTGAGAGAGTGTTTGGACAACAGAGTGGGTGTTAATGGCT
                                                                     350
  E  A  A  K  R  G  N  L  S  W  L  R  E  C  L  D  N  R  V  G  V  N  G

TAGACAAAGCTGGAAGCACTGCCTTATACTGGGCTTGCCACGGGGGCCACAAAGATATAGTGGAAATGCT
                                                                     420
  L  D  K  A  G  S  T  A  L  Y  W  A  C  H  G  G  H  K  D  I  V  E  M  L

ATTTACTCAACCAAATATTGAACTGAACCAGCAGAACAAGTTGGGAGATACAGCTTTGCATGCTGCTGCC
                                                                     490
  F  T  Q  P  N  I  E  L  N  Q  D  N  K  L  G  D  T  A  L  H  A  A  A
```

FIG.66A

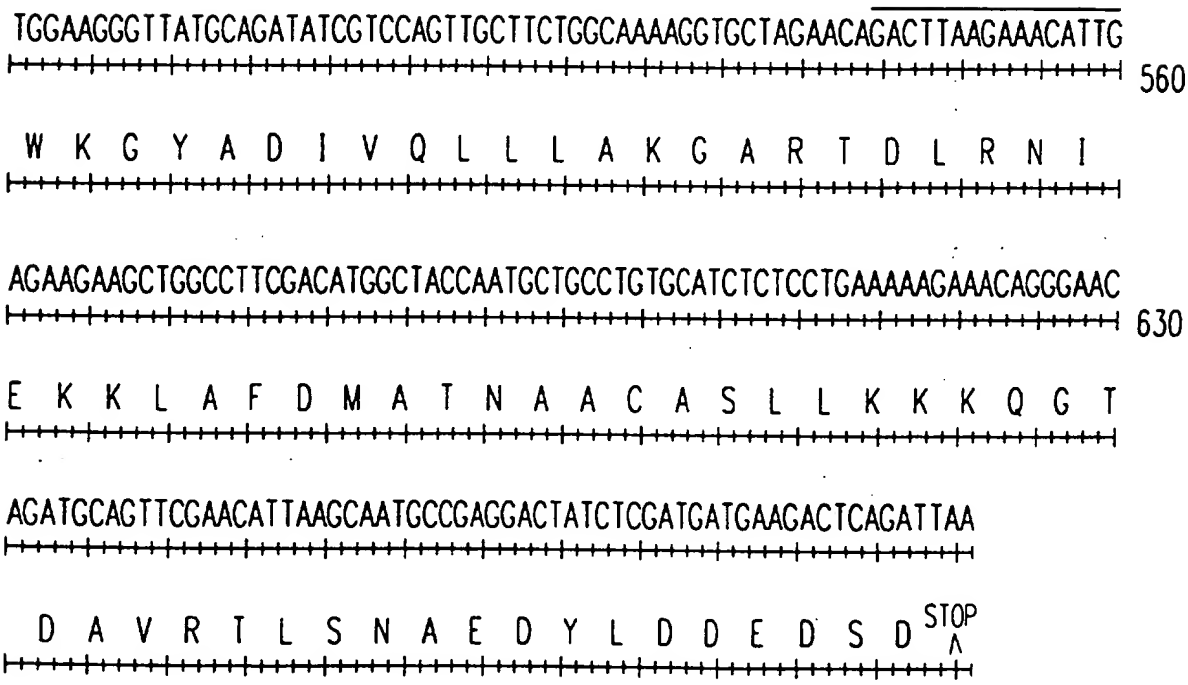
TGGAAGGGTTATGCAGATATCGTCCAGTTGCTTCTGGCAAAAGGTGCTAGAACAGACTTAAGAAACATTG
560
W K G Y A D I V Q L L L A K G A R T D L R N I
AGAAGAAGCTGGCCTTCGACATGGCTACCAATGCTGCCTGTGCATCTCTCCTGAAAAAGAAACAGGGAAC
630
E K K L A F D M A T N A A C A S L L K K K Q G T
AGATGCAGTTCGAACATTAAGCAATGCCGAGGACTATCTCGATGATGAAGACTCAGATTAA
D A V R T L S N A E D Y L D D E D S D STOP

FIG.66B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/04454

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12N 5/10, 15/12; C07K 14/00, 16/18; G01N 33/53
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 7.5, 172.1, 240.1, 320.1; 530/300, 350, 387.9; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	SPARKS et al. Cloning of ligand targets: Systematic isolation of SH3 domain-containing proteins. Nature Biotechnology. June 1996, Vol. 14, pages 741-744, see entire article.	1-102
X, P ---- Y	WO 95/24419 (ARIAD PHARMACEUTICALS, INC.) 14 September 1995, see pages 2-4, 13, 14, 16-18, 27 and 31, especially page 14, lines 16-22	53-95, 97-102 --------- 1-52, 96
X --- Y	SPARKS et al. Identification and Characterization of Src SH3 Ligands from Phage-displayed Random Peptide Libraries. The Journal of Biological Chemistry. September 1994, Vol. 269, No. 39, pages 23853-23856, see the abstract and page 23855, column 2.	53-95, 97, 11-102 --------- 1-52, 96

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 JUNE 1996	15 JUL 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer LORA M. GREEN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/04454

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	RICKLES et al. Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. The EMBO Jounal. 1994, Vol. 12, No. 23, pp5598-5604, see abstract and page 5602, column 2.	53-95, 107-102 --- 1-52, 96
Y	WO 93/18054 (N.V. INNOGENETICS) 16 September 1993, see pages 5, 24 and 25	1-52, 96

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/04454

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/04454

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/6, 7.1, 7.5, 172.1, 240.1, 320.1; 530/300, 350, 387.9; 536/23.5

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN, DIALOG
search terms: library, gene expression, peptide, avidin, biotin, multiple antigen peptide, phage display, antibody, SH3, SH2, zinc finger, leucine zipper

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.
Group I, claim(s) 1-52, 69-73, 89, 90 and 94-97 drawn to methods of identifying a polypeptide comprising a functional domain of interest.

Group II, claim(s) 53-68, 74, 75, 79, 80 and 101-102, drawn to a purified polypeptide, kits containing said purified polypeptide and methods of screening for a potential drug candidate.

Group III, claim(s) 76-78, 81-88 and 100 drawn to DNA encoding a polypeptide, a vector comprising said DNA, a recombinant cell and methods of producing a fusion protein.

Group IV, claim(s) 91-93, drawn to a method of determining the potential pharmacological activities of a molecule.

Group V, claim(s) 98 and 99, drawn to an antibody.

The inventions listed as Groups I-V do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The invention of Group I is drawn to a method of identifying a polypeptide comprising a functional domain of interest, and as claimed, does not require the products of Groups II, III and V. The polypeptide and kits of Group II have a defined seq. ID, which are not required in the method of Group I. In addition, functional domains such as SH3 domains are known in the art (see for Example, Cheadle et al., J. Biol. Chem. Vol. 269, No. 39, pages 24034-24039 (1994)). Further, the method of Group IV also lacks the technical feature of Group I as Group IV does not require the use of a multivalent recognition unit complex.
The invention of Group IV also does not have the same technical features as Groups II, III and V, as the method of Group IV as claimed does not require the products of Groups II, III and V. The polypeptide and kits of Group II have a defined seq. ID, which are not required in the method of Group IV. In addition, as stated above, functional domains such as SH3 domains are known in the art.

Groups II and III also lack a single concept. Group II is drawn to polypeptide and Group III is drawn to DNA, and thus have different structure and function. In addition, as stated above, polypeptides comprising functional domains such as SH3 domains are known in the art. Group V also does not relate to a single inventive concept, as Group V is drawn to an antibody, and is not required by the method of Groups I or IV, and is a separate product than the products of Groups II and III, having a different function and structure.